Chapter Five

Screening and Testing

Table of Contents

Figures

Tables

Appendices

I. Chapter Overview

This chapter describes the EDSTAC recommendations regarding development of a screening and testing program for assessing the potential of pesticides and other chemicals to disrupt endocrine function in humans and wildlife. Where appropriate, strengths and limitations of options are discussed and possible future research projects are identified to develop needed procedures. The EDSTAC established the Screening and Testing Work Group (STWG) (see Appendix D for a list of work group members) to assist in their efforts to provide guidance to EPA regarding the development and implementation of its endocrine disruptor screening and testing program. The STWG work formed the basis for this chapter and recommendations. Literature cited for all sections are found at the end of the chapter. Additional sources can be found in Appendix J.

After this introduction, the chapter is comprised of seven main sections: (1) the concepts and design parameters involved with Tier 1 Screening (T1S); (2) the Recommended T1S Battery; (3) the general principles in evaluating Tier 1 and Tier 2 results; (4) the concepts and design parameters involved with Tier 2 Testing (T2T); (5) the recommended T2T Battery; (6) a discussion of validation, standardization, methods development, and research; and (7) a summary of the recommendations made throughout the chapter.

The T1S sections begin with an explanation of the purpose of screening and identification of five criteria used to design the screening battery. An outline of the recommended T1S battery follows with brief overviews of each recommended assay and discussions of the value of including both *in vitro* and *in vivo* assays. Four alternative assays for consideration are also discussed. Finally, a section on evaluating the battery includes a discussion of a "weight-of-evidence" approach to evaluating T1S results.

In developing the T1S battery, the EDSTAC considered screening endpoints for their utility in screening chemical substances or mixtures for their potential to interact with the endocrine system. The goal of T1S is to detect chemical substances or mixtures capable of interacting with estrogen, androgen, or thyroid (EAT) hormone systems. Assessing these activities is relevant as changes in them may adversely affect the development, reproductive function, or chronic health status of humans or animals. The objective of T1S is not to determine dose-response relationships, confirm the mechanism of action, or determine the adversity of the chemicals' effect on reproduction and/or development; however, screening assays must be sensitive enough to detect all known xenobiotics that act via the mechanism of action each assay is designed to detect.

The screening battery presented here has been designed to ensure that interaction with hormone systems will be detected. There are instances in which a choice had to be made between an assay that was highly specific for a hormonal activity and one that may be less specific but more sensitive and apical (i.e., a more comprehensive assessment of functions that are relevant to reproduction, development, or chronic health). In those instances, the EDSTAC opted for the latter since it better fulfills the first criterion for screens (that they be sensitive), and is better aligned with the overall mission of detecting effects regardless of mechanism of action. These

assays require varying levels of additional development, standardization, and validation before they can be reliably and routinely implemented as part of T1S.

The T2T sections begin with discussions of the purpose of testing, guidance for selecting Tier 2 tests, and the issues of low dose considerations and selecting target doses for components of T2T. An outline of the recommended T2T battery is followed by overviews of the mammalian two-generation reproductive toxicity study, alternative mammalian tests, and tests using other vertebrate and invertebrate taxa. Finally, recommendations regarding implementation of the standardization, validation, and research program are presented.

The goal of T2T is to determine whether a chemical substance or mixture causes endocrine-mediated adverse effects for EAT and to determine the consequences to the organism of the activities observed in T1S and their dose response relationships. This is done in the larger context of testing for reproductive and developmental toxicity potential by any mechanism (including EAT) using study designs that provide a comprehensive assessment of relevant functions.

II. Tier 1 Screening Concepts and Design Parameters

Chemical substances or mixtures can alter endocrine function by affecting the availability of a hormone to the target tissue, and/or affecting the cellular response to the hormone. Mechanisms regulating hormone availability to a responsive cell are complex and include hormone synthesis, serum binding, metabolism, cellular uptake (e.g., thyroid), and neuroendocrine control of the overall function of an endocrine axis. Mechanisms regulating cellular response to hormones are likewise complex and are tissue specific. Because the role of receptors is often crucial to cellular responsiveness, specific nuclear receptor binding assays are included. In addition, tissue responses that are particularly sensitive and specific to a hormone are included as endpoints for Tier 1 screens.

The following definitions are utilized in this chapter. Estrogenic refers to compounds whose effects are mediated through the estrogen receptor (ER), initiating a cascade of cell/tissue specific effects similar to those initiated by estradiol, as opposed to estrogen-like for those chemicals resembling estrogen which are not or have not been shown to be mediated through the ER. Similarly, androgenic effects are androgen receptor (AR) mediated, as opposed to androgen-like effects, which may not be mediated via the AR. In contrast, the terms antiandrogenic and antiestrogenic are not specifically limited to AR- and ER-mediated interactions. In this context, agonists bind to the receptor and act like the endogenous hormone; antagonists bind to the receptor and appear to act opposite to the endogenous hormone. Antihormones can act via: (1) the steroid hormone receptor; (2) steroid hormone synthesis inhibition; (3) reduction of bioavailability by reducing the amount of free hormone in the serum; (4) increased hormone metabolism leading to reduced serum hormone levels; and (5) other mechanisms.

A. Introduction to T1S

The number of chemicals needing evaluation is huge. T1S is intended to make the evaluation process more efficient by distinguishing those chemical substances and mixtures which may interact with the endocrine system from those that may not. The EDSTAC considered all known endocrine disruptors of EAT in developing the T1S battery and believes that the recommended battery, if validated, will have the necessary breadth, and depth to detect all currently known disruptors of EAT. Therefore, following application of the T1S battery, a chemical substances or mixture will be designated as having either: (1) the potential for EAT activity, which will require further analysis in T2T to verify and evaluate that potential; or (2) low or no potential for EAT activity, which will allow assignment of chemical substances or mixtures to the "hold box" (see Section IV in this chapter, on general principles in evaluating Tier 1 and Tier 2 results for further discussion of how this decision is made).

In developing the recommended T1S battery, many existing and potential assays were evaluated for their relative strengths and weaknesses (overviews of evaluated assays can be found in Appendix K). The recommended T1S battery contains mammalian *in vitro* and *in vivo* assays and *in vivo* nonmammalian assays. The T1S battery is designed to be a sufficiently sensitive screening mechanism so that chemical substances and mixtures which ultimately prove to be endocrine active for EAT in humans and wildlife are not missed. In this regard, sensitivity of the battery has been prioritized above specificity. In addition, T1S results should inform T2T, in terms of providing guidance on which tests to perform, which endpoints to include, and to assist in determining the range of doses to be used. These goals include identifying those doses (dose-response), life stages (most sensitive), and organisms (most appropriate, sensitive, and at risk) in which adverse effects are likely to occur.

B. Criteria for T1S

The T1S battery recommended by the EDSTAC has been developed such that, at the completion of the selected assays, the EPA and other stakeholders will accept, both scientifically and as a matter of policy, the assignment of chemical substances or mixtures as either having: (1) <u>low or no</u> potential for estrogen, androgen, or thyroid endocrine activity; or (2) as having such potential. The ability to accept either outcome requires that the chosen T1S battery meets the five criteria identified below.

1. The T1S battery should maximize sensitivity to minimize false negatives while permitting an as of yet undetermined, but acceptable, level of false positives. This criterion expresses the need to "cast the screening net widely" in order not to miss potential EAT active materials.

2. The T1S battery should include a range of organisms representing known or anticipated differences in metabolic activity. The battery should include assays from representative vertebrate classes to reduce the likelihood that important pathways for metabolic activation or detoxification of parent chemical substances or mixtures are not overlooked.

3. The T1S battery should be designed to detect all known modes of action for the endocrine endpoints of concern. All chemicals known to affect the action of EAT hormones should be detected.

4. The T1S battery should include a sufficient range of taxonomic groups among the test organisms. There are known differences in endogenous ligands, receptors, and response elements among taxa that may affect endocrine activity of chemical substances or mixtures.

5. The T1S battery should incorporate sufficient diversity among the endpoints and assays to reach conclusions based on "weight-of-evidence" considerations. Decisions based on the battery results will require weighing the data from several assays.

The T1S must be relatively fast and efficient while meeting the criteria described above. The EDSTAC recommends that if changes are made to the recommended T1S battery, based upon development of new, validated assays, the "amended" battery also needs to meet these criteria.

III. Recommended Tier 1 Screening Battery

A. Outline of Recommended T1S Battery and Possible Alternatives[1]

1. Recommended T1S Battery

The T1S battery recommended by the EDSTAC includes three *in vitro* assays, three *in vivo* mammalian assays, and two *in vivo* nonmammalian assays. Those chemicals which go through the HTPS program, if it is technically feasible and validated, would not be required to do the first two *in vitro* assays at the bench. Based on existing data, the EDSTAC believes this battery will detect EAT activity, provided all of the component assays can be properly developed, standardized, and validated.

In Vitro

1. Estrogen Receptor (ER) Binding/Transcriptional Activation Assay;
2. Androgen Receptor (AR) Binding/Transcriptional Activation Assay; and
3. Steroidogenesis Assay with Minced Testis.

In Vivo

1. Rodent 3-Day Uterotrophic Assay (Subcutaneous);
2. Rodent 20-Day Pubertal Female Assay with Thyroid;
3. Rodent 5-7-Day Hershberger Assay;
4. Frog Metamorphosis Assay; and

[1] Protocols for all these assays can be found in Appendix L.

5. Fish Gonadal Recrudescence Assay.

2. Alternative Assays for Possible Inclusion

In addition, the EDSTAC has identified one *in vitro* assay and three *in vivo* assays as possible substitutes, if properly developed, standardized, and validated, for some of the component assays in the recommended battery. These assays are:

In Vitro

1. Placental Aromatase Assay.

In Vivo

1. Modified Rodent 3-Day Uterotrophic Assay (Intraperitoneal);
2. Rodent 14-Day Intact Adult Male Assay With Thyroid; and
3. Rodent 20-Day Thyroid/Pubertal Male Assay.

Combinations of the alternative assays, if validated and found to be functionally equivalent, could potentially replace three of the component assays in the recommended T1S battery (*in vitro* steroidogenesis assay with testis, 20-day pubertal female assay, and 5-7-day Hershberger assay), thereby possibly reducing the overall time, cost, and complexity while maintaining equivalent performance of the overall T1S battery. Table 5.1 shows the assays included in the proposed battery as well as two possible batteries that would include the alternative assays. In addition, Table 5.2 shows the assays in relation to which of the biological activities they are expected to detect, that may be affected by exogenous agents and lead to EAT-related toxicity.

One alternative battery would include the ER binding or transcriptional activation assay, the AR binding or transcriptional activation assay, the modified rodent 3-day uterotrophic assay (administered by intraperitoneal injection), the rodent 14-day intact adult male assay with thyroid, the frog metamorphosis assay, the fish gonadal recrudescence assay, and, possibly, the placental aromatase assay.

The other alternative battery would include the ER binding or transcriptional activation assay, the AR binding or transcriptional activation assay, placental aromatase assay, the rodent 3-day uterotrophic assay (administered by subcutaneous injection), the rodent 20-day thyroid/pubertal male assay, the frog metamorphosis assay, and the fish gonadal recrudescence assay.

Table 5.1

Assays Included in Recommended T1S Battery and Possible Alternatives

Assays	Recommended T1S Battery	Possible Alternative 1	Possible Alternative 2
In Vitro			
Estrogen receptor binding	x	x	x
Androgen receptor binding	x	x	x
Steroidogenesis	x		
Placental aromatase		?[1]	x
In Vivo			
3-day uterotrophic	x (sc) [2]	x (ip) [3]	x
20-day pubertal female	x		
Hershberger	x		
14-day intact male		x	
20-day pubertal male			x
Frog metamorphosis	x	x	x
Fish gonadal recrudescence	x	x	x

1 = may be needed in battery to meet criteria
2 = subcutaneous
3 = intraperitoneal

Table 5.2

T1S Assays Related to Biological Activities Detected

		Anticipated to Detect:								
Assay	in Option	Estrogen Agonism	Estrogen Antagonism	Androgen Agonism	Androgen Antagonism	Thyroid-Related Effects	Steroid Synthesis	Aromatase Inhibition	5-α-Reductase Inhibition	HPG[1]
In Vitro										
Estrogen receptor binding	1,2,3	X	X							
Androgen receptor binding	1,2,3			X	X					
Steroidogenesis	1						X			
Placental aromatase	3, 2?							X		
In vivo										
3-day uterotrophic	1,2,3	X		(X)[2]						
20 day pubertal female	1	X	X			X	X	X		X
Hershberger	1			X						LH[3]
Hershberger + T	1	(X)[4]			X				X	LH[3]
14-day intact male	2	X			X	X	X	(X)[5]	X	X
20-day pubertal male	3	X		X	X	X	X		X	X
Frog metamorphosis	1,2,3	X	?[6]	?[6]	?[6]	X	X	?[6]	?[6]	X
Fish gonadal recrudescence	1,2,3	X	X	X	X	?[6]	X	X	?[6]	X

Notes:

[1] HPG – indicates that the model has an intact hypothalamic-pituitary-gonadal axis (except for the Hershberger assay which does not have an intact HPG axis), and that effects on hypothalamic-pituitary control of gonadal endocrine function would be evaluated.

[2] It is likely that aromatizable androgens would be detected in this assay; however, given that there are no examples of environmental androgens, this point cannot be empirically demonstrated.

[3] Agents that affect LH level would be detected in the assay.

[4] Empirical demonstration that the assay detects estrogens is limited. The biology of the system suggests that they will be detected.

[5] Empirical demonstration that aromatase inhibitors are detected is limited. If sensitivity to aromatase inhibitors is lacking, a placental aromatase assay would be added to this option.

[6] The biology of these organisms suggests that these effects may be detectable. However, there are no empirical data to support the sensitivity of the assay for these endpoints.

3. Validation of the Battery

In order to provide sufficient data to allow informed decisions about the relative merits of the recommended T1S battery component assays and alternative assays (based on sensitivity, specificity, technical complexity, inter- and intra-laboratory variability, time, and cost), EDSTAC recommends that validation studies be initiated on all of the assays in the recommended battery as well as the alternatives.

If the assays comprising either of these alternative combinations (see Table 5.1) are validated, the EDSTAC recommends performance of the alternative battery containing these assays using the same standard test substances recommended for validation of the initial battery, which were selected on the basis of predetermined criteria (see Chapter five, Section VII, G). Sufficient information could then be available to allow an informed choice between the recommended battery or a variation, including the alternative assays, as the preferred T1S battery. This approach would be most expedient in the event that one or more of the recommended battery assays cannot be properly standardized and validated, since information would be immediately available on the alternative assays. The EDSTAC believes this process provides a model for validation and incorporation of new assays, as they may be developed and proposed, into the T1S battery.

The EDSTAC believes it is critical to acknowledge that the state-of-the-science, with respect to assay development and species selection, is rapidly evolving, and bioassays are currently being developed that may offer distinct advantages over those assays and species presently recommended for use. This is particularly the case for selection of non-mammalian species currently recommended for use in *in vivo* assays. Specific bioassays and species should be selected on a performance-based approach. As improved bioassays and/or those utilizing more appropriate species are developed and validated, EDSTAC strongly encourages their use as assays for screening and/or testing. Selected assays identified as research priorities by the EDSTAC are discussed in Section VII, F of this chapter. The EDSTAC recommends that EPA set up a specific mechanism for evaluating and incorporating these and other new developments, as appropriate, into the program.

Given the wide range of species that may be adversely affected by endocrine disruptors, continued development of screens and tests is particularly important to ensure that a representative range of species and potential endocrine-related effects can be evaluated.

4. Assays not Included in T1S

Currently, there are no data available to suggest that thyroid effects of chemical substances or mixtures are mediated through the receptor. Therefore, the recommended T1S battery does not currently include a thyroid receptor (TR) binding and/or transcriptional activation assay. Nevertheless, the EDSTAC is recommending that the HTPS program include evaluation of the TR. The Committee believes including the thyroid assays in the HTPS program will enable EPA,

and others, to obtain a better understanding, at relatively low costs, of whether effects could be mediated through the TR.

Brief overviews of all assays considered by the STWG can be found in Appendix K. In addition, Appendix M includes more thorough discussions of assays that the work group considered in detail, but decided not to include in the recommended T1S battery. The EDSTAC's thoughts on the role of invertebrates in T1S can be found in Appendix N.

5. Developmental Exposure Screening Assay

The EDSTAC considered all known endocrine disruptors of EAT in developing the T1S battery and believes, to the best of its knowledge, that the recommended battery, if validated, will have the necessary breadth and depth to detect any currently known disruptors of EAT. There is a concern, however, that chemical substances or mixtures could produce effects from prenatal/prehatch exposure that would not be detected from pubertal or adult exposure. Furthermore, there are differing views within the EDSTAC about whether there is scientific evidence of known endocrine disruptors or reproductive toxicants that can affect the prenatal stage of development without affecting the adult or prematuration stages, and whether effective doses and affected endpoints may differ among the three life stages.

Notwithstanding these differing views, the EDSTAC recommends that EPA take affirmative steps, in collaboration with industry and other interested parties, to attempt to develop a protocol for a full life cycle (i.e., with embryonic exposure and evaluation of the adult offspring) developmental exposure screening assay that can be subjected to validation and standardization. In addition to the general principles and criteria, set forth in Chapter Three, that guide selection of all T1S assays, the EDSTAC believes such an assay or assays must involve prenatal or prehatch exposure and retention of offspring through puberty to adulthood and structural, functional, and reproductive assessment.

The EDSTAC recognizes it may be difficult to develop a developmental exposure screening assay that meets both the criteria specified above, and the more general criteria for selecting T1S assays set forth in Chapter Three. However, the EDSTAC believes it is worth the effort. Furthermore, in Section VII, F, the EDSTAC has summarized protocols for *in utero* and *in ovo* developmental screening assays that could be further evaluated for this purpose (an expanded discussion of an *in utero* protocol is included in Appendix O). Inclusion of these protocols is not intended to limit the creative effort that will be necessary to achieve the EDSTAC's recommendation.

Finally, the EDSTAC recommends that if such an assay were identified, validated, and standardized, the decision on whether it should be included in the T1S battery should include an evaluation of its potential to replace one or more of the recommended T1S assays and its overall impact on the cost effectiveness of the T1S battery. It should be noted, however, full life cycle assessments are included in the recommended T2T battery for mammals, other vertebrates, and invertebrates. These tests will employ a full range of doses, embryonic exposures, rearing offspring to adulthood, and a full complement of reproductive and developmental endpoints.

6. Methods to Select the Appropriate Dose Level(s) for *In Vivo* Assays

All T1S *in vitro* assays (including the steroidogenesis assay) will involve multiple dose levels, whether performed by HTPS or bench level methods, so a dose-response curve and assessment of relative potencies can be developed. Results from the HTPS (or its equivalent) will provide potency information (i.e., EC 50) relative to a positive control such as 17 beta estradiol (E2), diethylstilbestrol (DES), testosterone, or T4 for those chemical substances or mixtures which bind to the E, A, or T receptor. Information on the *in vitro* effective doses of E2, DES, testosterone, or T4 can be used to set the dose level(s), based on the validation process, for the remaining T1S assays for these chemical substances or mixtures. There are no current data which indicate that thyroid toxicants act via binding to the thyroid hormone receptor(s). Thus, the recommended *in vitro* receptor binding/transcriptional activation assay may not inform dose selection for *in vivo* T1S assays for thyroid endpoints. For these chemical substances or mixtures, prior information and range-finding studies will be critical.

Subject to the results of the validation process, the EDSTAC recommends using one or more dose levels in the performance of the *in vivo* assays. Information to assist in selecting the dose level(s) includes:

1. prior information, such as that available during the priority setting phase;
2. results from the HTPS (or its equivalent bench-level assays); and
3. results from range-finding studies, utilized for T1S dose selection (see below).

A range-finding study can be performed at multiple dose levels (at least five) with a few animals per dose level and a limited number of relevant endpoints. The range-finding studies specifically performed for each *in vivo* TIS assay will include the following:

- use of the same species strain, sex(es), and age as in the T1S assay;
- use of the same route of administration, vehicle, and duration of dosing as in the T1S assay;
- use of multiple dose levels (the number of dose levels will depend on the availability and extent of prior information);
- use of multiple animals per dose level which may be fewer than the number used per group in the T1S assay;
- use of relevant endpoints, which may be more limited than those in the T1S assay (for example, the range-finding study for the T1S uterotrophic assay may employ only body weights and uterine wet weight, while the assay may also evaluate uterine gland height, serum hormone levels, and/or vaginal cornification, etc.);
- use of comparable animals (e.g., ovariectomized females for the uterotrophic range-finding study or castrated males for the Hershberger range-finding assay). However, there may be circumstances under which exceptions occur (e.g., use of intact males in the range-finding study for the Hershberger assay to define doses producing systemic toxicity and any effects on the reproductive system as a first pass approximation); and
- use of more than one range-finding study if the initial version does not identify the dose level(s) to be used in the specific T1S assay if necessary by extrapolation or interpolation.

The dose level(s) to be selected for the *in vivo* assays should not result in excessive systemic toxicity, but should result in effects useful for detection of potential EAT disruption. However, no dose level higher than one gram/kilogram body weight/day (i.e., a "limit" dose) should be utilized. The rationale for selection of dose levels for each range-finding study, all of the results for such studies, and the logic employed to select the dose level(s) for the T1S assay should be included in the submission of T1S results for evaluation by the Agency as to the appropriateness of the study design, conduct, and conclusions.

7. Routes of Administration

The route of administration for the recommended uterotrophic assay is subcutaneous (sc) injection while the route for the modified uterotrophic assay and 14-day intact adult male assay with thyroid is intraperitoneal (ip) injection. The route for all other mammalian *in vivo* assays is gavage (orogastric intubation). The parenteral (non-oral) routes avoid the first-pass metabolic effect of the liver and will permit detection of potential EDCs that are active as parent compounds and which undergo significant first-pass metabolism. Hepatic xenobiotic metabolism does occur eventually after parenteral administration (substantially with ip), so the potential effects of metabolites will be evaluated as well by these routes. Compounds are occasionally metabolized by the gut microflora; this type of metabolism has been shown to be important for some plant-derived estrogens. The oral route of exposure will allow for this type of metabolism.

The EDSTAC believes EPA should propose a policy for route of administration for each assay and test. Since T1S seeks activity, an ip route or other *in vivo* approach seems most realistic for detecting potential endocrine activity. Conversely, T2T should focus on developing a policy for route of administration based on exposure route(s) which approximates the ecologically relevant exposure pathway, dependent on the test species and fate of the chemical in the environment.

B. Scientific Basis for *In Vitro* Screening for Estrogen, Androgen, and Thyroid Activities

General agreement has been reached on the strengths and limitations of most currently available *in vitro, in vivo,* and *ex vivo* methods for detection of toxicants that act via ER, AR, steroid hormone synthesis inhibition, and/or altered hypothalamic-pituitary-gonadal (HPG) mechanisms. With this in mind, several short-term *in vitro* assays for AR and ER receptor binding or transcriptional activation and minced testis steroidogenesis inhibition (SI) activity were identified as quite useful in screening. *In vitro* methods also include steroidogenic enzyme/hormone synthesis, biochemical assays, and *in vitro* and testis steroid hormone synthesis.

Advantages of *in vitro* assays include*:*

a) sensitivity to low concentrations increases detectability;
b) high specificity of response;
c) low cost;
d) small amount of chemical substance or mixture required;

e) *in vitro* assays can be automated, including use of robotics;
f) high throughput assays (thousands/month) can be developed;
g) results can be coupled with QSAR models and for database screening;
h) can be used for complex mixtures (sludge, water contaminants); and
i) reduces or replaces animal use.

The EDSTAC recognizes two categories of *in vitro* assays that may be used in T1S to assess the binding of test substances to receptors, i.e., cell-free assays for receptor binding and transfected cells designed to detect transcriptional activation. The specific assays chosen, whether done "at the bench" or through the high throughput pre-screening process (discussed in detail in Chapter Four, Section V), should have the following characteristics:

a) evaluate binding to estrogen, androgen, and perhaps thyroid nuclear receptors;
b) evaluate binding to the receptor in the presence and absence of metabolic capability (e.g., one or more of the P450 isozymes, CYP1A1, CYP3A4, etc.);
c) distinguish between agonist and antagonist in functional assays; and
d) yield dose responses for relative potency of chemical substances or mixtures with endocrine activity.

If high throughput procedures are used, receptor binding assays should be performed for EAT receptors. If the assays are done at the bench level, only estrogen and androgen receptor assays are recommended and/or functional assays should be performed for estrogen, androgen, and perhaps thyroid receptors (specifically recommended is a stably transfected cell line like the MVLN cell line, if available, to assess transcriptional activation). If stably transfected cell lines are not available, then transiently transfected reporter gene assays should be used. MCF-7 proliferation assays are also acceptable; however, yeast-based assays are not recommended at this time. These assays can be performed either high throughput or at the bench level.

Receptor binding assays can use rat, mouse, or human ER or AR. These assays evaluate the ability of the xenobiotic chemical substances or mixtures to displace the radio-labeled endogenous ligand from the binding site, in a cell-free or whole cell system. Relative potency can be determined for positive chemical substances or mixtures. Assay limitations are solubility in the culture medium, inability to distinguish agonists from antagonists, lack of metabolic capability, and risk of degradation of the receptor.

The functional assay, specifically transcriptional activation, requires, for agonist or antagonist activity, that the chemical substance or mixture bind to the receptor. In addition, there is a consequence to the binding, i.e., transcription (synthesis of mRNA) of a reporter gene and translation of the mRNA to an identifiable detectable protein such as firefly luciferase or beta-galactosidase. In the case of the firefly luciferase, with substrate and cofactors present in the culture, there is a light flash detected from formation of the product when the enzyme is synthesized in response to transcriptional activation and acts on the provided substrate. In the case of the beta-galactosidase, with substrate and cofactors present in culture, the product is detected colorimetrically when the enzyme is synthesized in response to transcriptional activation and acts on the provided substrate. The assay uses intact cells and may use different cell lines for

assessment of effects on EAT binding domains with transfected (transiently or permanently) receptors and reporter gene constructs. This assay can distinguish between agonists and antagonists. Assay limitations are solubility, toxicity, permeability of the cell membrane, and lack of or limited metabolic capability. If a chemical substance or mixture must be metabolized to an active moiety, it will not be detected unless the limited residual metabolic capacity of the cultured cells is sufficient to transform the chemical to its active form. Metabolic activity might be provided by either preincubating the chemical substance or mixture with an S9 fraction (supernatant from 9000g x centrifugation of homogenized liver from a metabolically induced rat) or incorporating the S9 fraction into the treatment mixture. In addition, cell lines are being genetically engineered to incorporate genes for P450 enzymes as a method for extending their metabolic capacity and, perhaps, obviate the need for use of the S9 fraction.

For assessing receptor binding *in vitro*, EDSTAC recommends both the receptor binding assays and the transcriptional activation assays be incorporated into the T1S battery, and subjected to validation and standardization. There is agreement that the transcriptional activation assays can provide more information than the receptor binding assays, since they measure not just binding capacity but also the physiological and biochemical consequences of that binding. However, the limited database on the relative utilities of receptor binding and transcriptional activation assays do not allow the EDSTAC to recommend one category of assay over the other at this time. Including the receptor binding and transcriptional activation assays in the standardization and validation program is expected to provide the data needed to reach a decision on whether both assays should be required or, if not, whether the receptor binding or transcriptional activation is preferred. It is important to keep in mind that these assays evaluate just one of the possible mechanisms of endocrine disruption; if a chemical substance or mixture acts via another mechanism than the receptor, it will not be detected in these assays.

Large-scale high throughput pre-screening (HTPS) programs for chemicals have been employed, using standardized *in vitro* functional assays (i.e., transcriptional activation of a reporter gene), in the pharmaceutical industry. Several companies involved in drug design routinely screen chemicals for hormonal activity on a large scale (thousands per month).

In vitro evaluations can provide both false positive and false negative results. *In vitro* false positives (i.e., active *in vitro* but not *in vivo*) arise when a chemical is not absorbed or distributed to the target tissue, is rapidly metabolically inactivated and excreted, and/or when some other form of toxicity predominates *in vivo*. False negatives are considered to be of greater concern if *in vitro* tests were used to the exclusion of *in vivo* methods. *In vitro* evaluations can result in false negatives due to their inability, or unknown capacity, to metabolically activate toxicants. As a result, the EDSTAC's recommended battery includes *in vivo* methods in conjunction with *in vitro* techniques. Nevertheless, some *in vitro* assays may offer distinct advantages over *in vivo* assays when investigating the activity of specific metabolites.

C. *In Vitro* Assay Overviews

The EDSTAC recommends a specific assay for each of the ER receptor binding, ER transcriptional activation, AR receptor binding, AR transcriptional activation, and steroidogenesis categories in order for standardization and validation to occur efficiently. The receptor binding and transcriptional activation assays would be performed only on those chemical substances or mixtures not going through HTPS, while the steroidogenesis assay would be performed on all chemical substances or mixtures going through T1S. Equivalent assays could replace these if they meet specific performance criteria and were similarly validated. Even if HTPS is implemented, standardization and validation of these additional *in vitro* assays would allow them to be conducted in individual labs on a more limited basis. The following assays are the specific ones recommended for inclusion in the standardization and validation program.

1. Estrogen Receptor Assays
   a) ER Binding: Cell-Free ER Alpha Binding
   b) ER Transcriptional Activation: MVLN

2. AR Assays
   a) AR Binding: Cell-Free AR Binding
   b) AR Transcriptional Activation: AR Transcriptional Activation

3. Steroidogenesis
   a) Minced testis

1. Estrogen Receptor Assays

In vitro rat ER binding assays provide a rapid and fairly inexpensive method for quantifying the ability of chemicals to compete with DES or estradiol for ER. The assay can be used for measuring ER in cell-free extracts obtained from various tissue homogenates following *in vivo* exposure to an environmental chemical. In addition, the assay may be used to determine the ability of a given compound to compete with radio-labeled estradiol for binding to the ER. The technical aspects of the ER binding assay are well documented for receptors obtained from cytosolic or nuclear extracts of various mammalian and other vertebrate tissues (Anderson et al., l972; Korach et al., l979). In brief, cytosolic or nuclear extracts containing ER are incubated with [^{3}H] estradiol for 18 hours at 4^o C in the presence or absence of increasing concentrations of radio-inert DES or test chemicals. Nonspecific binding is assessed by the addition of 100 molar excesses of radio-inert DES. Bound [^{3}H]-and free ligands are separated using hydroxyapatite extraction, or charcoal-dextran adsorption, and are quantified by scintillation counting.

The ER binding assays are less sensitive than the functional assays, of short-term duration, and can be standardized between laboratories. The assay is useful for evaluating effects of a test compound on ER distribution and number following *in vivo* exposure. In addition, the assay can be used to rapidly evaluate test compounds for their ability to bind to the ER in the absence of any of their metabolites. Comparison of IC_{50} and K_i values for the chemicals tested *in vitro* with that of endogenous and synthetic estrogens provide an indication of the potential of a given chemical to disrupt ER function *in vivo*. However, this assay does not distinguish between ER agonist and antagonists. The cytosolic rat ER binding assay may also yield false negative results if metabolic activation is required prior to binding

to the ER or if the test chemical is not completely solubilized in the assay buffer. In addition, the results may be artifactual if ER is altered by detergent/denaturation effects of the test chemical, particularly if concentrations greater than 10 micromolar are used. At present, ER binding data are not entirely comparable from lab to lab because of methodological differences between labs in the conduct of this assay. However the rat cytosolic ER binding assay has been used for about 20 years; it is less complex than whole cell binding assays, and competent laboratories should be able to obtain similar results with minimal effort.

Cell-free and whole-cell binding assays using human ER (hER) are rapidly being developed and offer both advantages and disadvantages over the above assay, one advantage being the use of the human rather than the rat ER. However, being relative new, they have not been standardized in their examination of xenoestrogens. Assays for ER beta binding and/or transcriptional activation should be considered as they become more widely available, and included in screening if warranted (i.e., if it is determined that some xenobiotics bind only to, or more avidly to, ER beta and would be missed in current assays with ER alpha).

a. ER Binding

The cell-free estrogen receptor alpha binding assay, a long-standing and relatively simple *in vitro* assay that detects specific mechanisms of endocrine activity, is recommended. This is important because several xenobiotics display affinity for the estrogen and/or androgen receptors. Binding assays identify, but do not discriminate between, agonists and antagonists. The apical nature of these assays is an advantage rather than a limitation because either activity can produce adverse reproductive effects. These assays typically lack metabolic activity, which is an advantage if one wishes to identify the specific compound with endocrine activity. However, the lack of metabolic activation is also a limitation because some xenobiotics require metabolic activation.

b. ER Transcriptional Activation

Binding of estrogen to ER alpha in target cells results in the initiation of specific transcription activation events. Various estrogen-regulated genes have been identified in MCF-7 cells (pS2, Cath D, PgR, TPA), and their corresponding gene products can be measured as an endpoint for estrogen action (VanderKuur et al., 1993a; Pilat et al., 1993; Davis et al., 1995). However, such endogenous genes are additionally regulated by other cellular mechanisms (Nunez et al., 1989; Cavailles et al., 1989; Zacharewski et al., 1994), and the quantification of gene products (mRNA) may be relatively laborious and difficult. Therefore, the introduction of artificial, ER-regulated reporter gene constructs into MCF-7 cells has become a routine method of measuring ER transcriptional activation (VanderKuur et al., 1993b; Meyer et al., 1994). These reporter assays utilize the human ER of MCF-7 cells for transcriptional regulation of a reporter gene that codes for an exogenous enzyme that can be easily measured in a cell lysate. Of the typical reporter gene products of chloramphenicol acetyl transferase (CAT) and luciferase (Luc), the more sensitive assays utilize luciferase. Reporter genes can be introduced into cells for the duration of the experiment only (transient transfection) or permanently, creating a genetically altered subline (stable transfection).

Transcriptional activation assays are a direct manifestation of receptor-mediated responses on gene expression (i.e., the presence of a functional estrogen receptor and a reporter gene are sufficient to express estrogen-mediated induction). The **MVLN** assay (stably transfected MCF-7 cell line with an artificial gene including ER alpha, a controller segment of vitellogenin, and promotor regulating expression of luciferase), which detects transcriptional activation after receptor binding using a luciferase reporter gene, is recommended. This rapid and sensitive assay (IC50=20 pM range) confirms ER binding and appropriate controls can distinguish agonists from antagonists. These assays should be conducted in a manner that allows them to detect receptor antagonists as well as agonists. Although these assays often provide information similar to the above binding assays, this is not always the case, and there are well-founded biological reasons for a chemical to be positive in either the binding or the transcriptional activation assay but not both. However, due to a higher degree of difficulty, concern exists that proper execution of whole-cell assays requires a level of skill and training that may not currently exist in the toxicology community. If so, these assays might be much more difficult to implement than the binding assays, some of which have been used for decades and are less complex.

In spite of the difficulty of establishing stably transfected cell lines, various MCF-7 cell derivatives have been created. As mentioned above, the MVLN cell line is an MCF-7 cell derivative containing an artificial gene consisting of the ER-controlled segment of the vitellogenin promoter, regulating the expression of luciferase (Pons et al., 1990; Gagne et al., 1994). These cells also contain a neomycin resistance gene that was used in the stable transfectant selection process. Therefore, since all MVLN cells contain the reporter gene, estrogen-regulated transcription can be measured with a high sensitivity. However, the metabolic capability of the MVLN assay has not been studied in detail; it is assumed to be similar to that of MCF-7 cells from which they are derived. In principle, there are several advantages of this assay over other *in vitro* assays that assess estrogen action. The MVLN cell assay is easy to use because it is permanently transfected and it is a short-term assay. In addition, the MVLN cell assay has been standardized to the degree that is has been employed in high throughput transcription assays involving robotic manipulation of large numbers of sample wells containing relatively few cells (e.g., 96-well plates). A procedure that has been used to characterize estrogen agonists as well as antagonists can be characterized with the MVLN assay (Gagne et al., 1994). In addition, a systematic comparison of more than 25 chemicals, including phthalates, alkylphenols, chlorinated pesticides, and steroids in the MVLN and the MCF-7 proliferation assay found that these assays were of equivalent sensitivity and responsiveness. Assays like the MVLN are deemed desirable because they are stably transfected and hence relatively easy to use and standardize, have high throughput potential, and are typically run to detect both agonists and antagonists.

The MVLN assay has been reported to have a disadvantage though, namely, that when the cells are briefly exposed to hydroxytamoxifen, their reporter gene cannot respond to estrogens. The mechanism underlying this effect is presently unknown. In principle, avoiding exposure to hydroxytamoxifen should prevent this from happening; however, this raises the issue of instability due to inadvertent exposure to chemicals during maintenance or propagation of the cells (this requires a serum-supplemented medium). The MVLN cells, like all other cell culture models, requires monitoring in order to ascertain that the initial response is preserved through extensive propagation (Badia et al., 1994). In addition to the MVLN, other stably transfected cell lines have been or are being used to detect for ER and AR action.

2. Androgen Receptor Assays

a. AR Binding

The **cell-free AR binding** assay, used to determine the ability of environmental chemicals to compete with endogenous ligand for binding to AR, is recommended. This is an easy, time-honored task, with decades of use, and relatively simple to standardize and execute. Equilibrium binding assays require overnight incubation at 4°C with AR isolated from castrated rat reproductive tissues (e.g., epididymis, ventral prostate, seminal vesicle) with increasing concentrations of radio-labeled ligand at different fixed concentrations of inhibitor or a fixed concentration of labeled androgen with increasing concentrations of unlabeled competitors. Following the incubation, hydroxyapatite or dextran-coated charcoal is used to separate protein-bound ligand from free ligand and specific binding is plotted in double reciprocal plots (i.e., Lineweaver-Burke) and as Scatchard plots as a function of competing inhibitor concentrations. Data analysis yields apparent equilibrium binding affinity constants for the inhibitor (Ki), which reflects the affinity of the chemical for the AR. Ki values can be used to rank chemicals for their ability to bind AR and therefore for their potential to be endocrine active. IC_{50} values can be used to calculate Ki values and the relative binding affinity (RBA) of the toxicant for AR, as compared to DHT or T, but this method is less accurate than experimental determination of the Ki. Within the last few years, a surprising number of chemicals in the environment of anthropogenic origin have been shown to act as AR ligands, including pesticides (e.g., vinclozolin, procymidone), pesticide metabolites (p,p' DDE and other DDT metabolites, methoxychlor metabolites), hydroxylated PCBs, and steroidal and non-steroidal natural and synthetic estrogens (Waller et al., 1996).

Advantages of the cell-free binding assay include ease of use, low cost, the potential to standardize receptor preparations for distribution to many labs, and metabolism (but not spontaneous degradation) of chemicals in the assay is minimized. The absence of metabolism is an important consideration as parent chemicals and/or metabolites can be individually examined to determine which structure is responsible for AR binding, information that is critical if the data are to be used in a QSAR model. Disadvantages include the need for radio-labeled ligands and that data are restricted only to ligand binding affinity with no information on agonist or antagonist activity, AR stabilization, or degradation or rates of association and dissociation from the AR.

b. AR Transcriptional Activation

For AR-mediated activity, stably transfected cell lines are under development, but not yet widely available. The **AR transcriptional activation** (Cis-Trans) assay, using monkey kidney CV-1 cells, is recommended. A MCF-7 cell stably transfected with wild type androgen receptor has recently become available; however, only a few androgen agonists and antagonists have been tested using this cell proliferation assay (Szelei et al., 1997). Hence, like the CV-1, cell lines transiently cotransfected with hAR and a promoter construct with a Luc reporter are recommended at this time. It is noteworthy that as compared to MCF-7 cells, the CV-1 has some metabolic capability. Here again, the YAS is not acceptable as it is unable to detect the AR-mediated activity of chlorinated pesticides.

Cells transiently transfected with hAR and reporter construct to detect transcriptional activation after receptor binding distinguish agonist/antagonist. Such assays have been used extensively and can be

employed in a HTS mode for rapid screening. Transcriptional activation assays are used to determine whether chemicals which bind AR act as AR agonists or antagonists (Zhou et al., 1994; Simental et al., 1991). CV-1 cells are transiently transfected with the hAR expression vector together with a reporter construct (e.g., chloramphenicol acetyl transferase (CAT), beta-galactosidase, or firefly luciferase) containing an AR-dependent promoter such as the mouse mammary tumor virus promoter. Transfected cells are cultured in the presence (for antagonist activity) or absence (for agonist activity) of a single concentration of androgen (0.1 nM DHT) together with increasing concentrations of inhibitor. Following a 48 hour culture period, cells are harvested and luciferase activity is measured in the resultant solubilized cell extract as an estimate of AR-induced transcriptional activity.

Advantages of these types of assays are that they use human AR, they display some metabolic activity, and they establish whether a chemical that binds hAR acts as an agonist or antagonist. This information is critical in understanding the mechanism responsible for the induction of adverse endocrine-mediated effects. Disadvantages of these assays are that they require the AR expression vector, reporter vectors, and transient cotransfections, which can be difficult. The assay requires close adherence to the standard operating procedure for reproducibility, and a 48 hour incubation during which time metabolism of the treatment chemicals may confound the data. In this regard, media from this assay, and other *in vitro* assays, should be analyzed before and after the incubation period to account for potential degradation and metabolism of the exogenous test chemicals and hormones.

3. Steroidogenesis

Antiandrogens and antiestrogens act via a number of direct mechanisms in addition to those that directly involve the steroid hormone receptors. One prominent mechanism of antihormonal activity is inhibition of hormone synthesis by inhibiting the activity of P450 enzymes in the steroid (and fungal) pathway. Such activity can be detected *in vitro* with a fairly simple *in vitro* procedure with minced testicular tissue obtained from adult male rats, because for many of the pesticides known to alter this pathway the parent material is active. Although aromatase, another P450 enzyme is present only at very low levels in the testis and male reproductive tract, it was proposed that inhibition of aromatase need not be included *in vitro* because it will be assessed in the *in vivo* pubertal female assay that follows. However, aromatase activity cannot be assessed in the recommended testis culture assay or in any of the *in vivo* assays using male rats.

The **testis culture *in vitro* assay using minced (50 mg) pieces of single testis,** which can be used to evaluate hormone synthesis with and without stimulation with cAMP, hCG, or substrates, is recommended. This assay assesses non-receptor mediated effects on P450 steroidogenic enzymes. Incomplete metabolism *in vitro* is of concern, except for those classes of chemicals where the parent material is active (e.g., certain classes of fungicides, drugs, and agricultural products). This assay has been used with fetal, neonatal, and adult testis, and is not limited to mammalian species, having been used to assess steroidogenesis in fish, reptile, avian, and amphibian systems as well.

It is also possible to use cultures of Leydig cells isolated from testicular tissue to perform steroidogenesis assays. Leydig cells are the cells, within the testis, responsible for steroid synthesis. The advantage of using these isolates is that they are enriched for the cells that synthesize testosterone. The disadvantage is that there are extra steps in the preparation of the

cells. Both approaches are expected to be comparable in their ability to detect steroidogenesis inhibitors. The utility of the minced testis culture is primarily based on data generated using Leydig cell cultures (Klinefelter and Kelce, 1996).

Substances that interfere with steroidogenesis act primarily by inhibiting cytochrome P450 enzymes in the steroid pathway. For example, the mechanism of action of two major classes of herbicides, the imidazoles and the triazoles, involves inhibition of P450 enzymes in the sterol synthesis pathway for lanosterol, a vital component of fungal membranes (Taton et al., 1988). Cytochrome P450 inhibitors tend to be non-specific, and these fungicides can also inhibit other P450 enzymes such as those required for mammalian steroid hormone synthesis (Murray and Reidy, 1990). Inhibition of mammalian steroid synthesis can potentially result in a broad spectrum of adverse effects *in vivo*, including abnormal serum hormone levels, pregnancy loss, delayed parturition, demasculinization of male pups, lack of normal male and female mating behavior, altered estrous cyclicity, and altered reproductive organ weights.

D. Scientific Basis for *In Vivo* Screening for Estrogen, Androgen, and Thyroid Activities

The EDSTAC believes inclusion of *in vivo* methods in T1S can help reduce false negatives in the absence of knowledge of absorption, distribution, metabolism, and excretion. *In vivo* assays are often apical (that is, while they incorporate endocrine-specific endpoints, disruption of a number of hormone regulation/delivery mechanisms can be evaluated in the same assay). Therefore, they are less specific, but more comprehensive, than *in vitro* assays. *In vivo* assays can be made more specific if accompanied by target organ/cell dosimetry of biologically active metabolites. *In vitro* data are enhanced if the actual concentration of the chemicals in the media is determined, to account for metabolism, stability, and solubility, and to determine whether these concentrations compare to those that can be achieved *in vivo*. Cellular assays should determine viability, and the specificity and limitations of each assay should be defined. It is clear a combination of *in vivo* and *in vitro* assays is necessary in order to detect EAT alterations that act via the ER, AR, TR, inhibition of steroid hormone synthesis, and/or alterations of the hypothalamic-pituitary-gonadal (HPG) and HPT (thyroid) axis.

More than 50 assays, and related endpoints, were considered by the STWG, including *in vitro*, *in vivo*, and *ex vivo* (*in vivo* dosing followed by *in vitro* assessment of function) techniques. *In vivo* endpoints considered include reproductive organ weights and histology, serum hormone levels, i*n vivo* gene activation, protein synthesis, behavior, growth, development, pregnancy maintenance, and anatomy/morphology. For each endpoint, the sensitivity (defined here as the response of the assay to low concentrations or dosage levels), specificity (pathognomonic for a mechanism of action, since the lack of specificity leads to false positives), relative simplicity, difficulties encountered running the assay, confounding factors, and limitations, test duration, and costs were discussed. In addition, items such as degree of acceptance of the method, how many chemicals had been screened, and the relative "newness" of the assay (state-of-the-art) were considered.

Advantages of *in vivo* assays include:

a) account for absorption, distribution, metabolism, and excretion;
b) well-defined, acceptable methods used for decades;
c) general acceptability in toxicity testing;
d) some endpoints are toxicologically relevant and have been used in risk assessment;
e) evaluate a broader range of mechanisms;
f) provide a comprehensive evaluation of the whole endocrine system as a unit; and
g) give comparative perspective to other endpoints of toxicity.

It is important to reiterate that the screening battery is being designed to minimize false negatives, based on an assessment of the ability of the battery to detect known EDCs that act via EAT. In this regard, the value of each individual assay cannot be considered in isolation from the other assays in the battery, as they have been combined in a manner such that limitations of one assay are complemented by strengths of another.

The EDSTAC believes the recommended screening battery, once validated, will detect all of the EDCs mediated by EAT including xeno(anti)estrogens (that act via the ER or inhibition of aromatase by oral or parenteral administration), xeno(anti)androgens (via AR or hormone synthesis), altered HPG axis, and antithyroid action (via synthesis, metabolism, and transport, and the TR). However, results of even the most specific *in vivo* assays can be affected by endocrine mechanisms other than those directly related to ER, AR, and TR action. For example, uterine weight in the ovariectomized female rat is affected in an estrogen-like manner by high doses of aromatizable and nonaromatizable androgens and growth factors like EGF. The age at puberty (vaginal opening in the female or preputial separation in the male rat) can be affected by chemicals that act on the hypothalamus, pituitary, or thyroid or alter growth hormone secretion. If gonadally intact females are used, uterine weight can also be affected by toxicants that stimulate hypothalamic-pituitary or gonadal endocrine secretions. Clearly, castration of the treated male or female markedly affects the specificity of the test. The lack of specificity of *in vivo* assays is a limitation if the goal is to only identify ER, AR, and TR alterations. In contrast, this lack of specificity could be considered an advantage if a broader, more apical screening strategy is desired.

1. Unique Thyroid Action Properties to be Considered in Design and Interpretation of T1S

Thyroid dysfunction leads to abnormal development, altered growth patterns, and a variety of physiological perturbations in mammals (Dussault and Ruel, 1987; Myant, 1971; Porterfield and Hendrich, 1993; Porterfield and Stein, 1994; Timiras and Nzekwe, 1989), as well as in birds (Tsai and Tsai, 1997), reptiles (Schrebier and Richardson, 1997), amphibians (Brown et al., 1995; Tata, 1994), and fish (Leatherland, 1994). Considering the consequences to wildlife populations and human health of the presence in the environment of synthetic compounds with thyroid disrupting activities, the EDSTAC has recommended a series of assays that will detect whether substances may interact with the thyroid.

The chemistry of thyroid hormone, the endocrine mechanisms governing its regulation, and the mechanisms by which thyroid hormone exerts its effects are surprisingly similar among vertebrates

(Gorbman et al., 1983). The EDSTAC deliberations have, therefore, been guided by research focused on a variety of vertebrates to develop this series of screens. Despite the volume of literature reviewed, the rapid pace of research into thyroid hormone action makes it predictable that the present screens will become obsolete, both because more effective assays will likely be developed and because new information about thyroid hormone action may reveal mechanisms of thyroid disruption not identified by the recommended T1S battery. The following background information, about functioning within the thyroid axis and methods used to evaluate anti-thyroid activities, is intended to provide a rationale for the recommended Tier 1 thyroid-specific assays.

Endocrinology of the Vertebrate Thyroid:

Cells of the thyroid gland are arranged in follicles; the epithelial cells surround a fluid-filled core containing proteinaceous material – the colloid (Fawcett, 1986). Individual follicular cells respond to a pituitary hormone, thyrotropin (TSH), by increasing the synthesis and release of thyroid hormones (Wondisford et al., 1996). In mammals, TSH release from the pituitary is stimulated, in turn, by a neuroendocrine peptide, thyrotropin-releasing hormone (TRH) (Greer et al., 1993; Morley, 1981; Taylor et al., 1990), and inhibited by the negative feedback effects of thyroid hormone itself (Franklin et al., 1987; Mirell et al., 1987; Shupnik and Ridgway, 1987). In a redundant negative-feedback loop, thyroid hormone also exerts an inhibitory effect on brain cells that manufacture TRH (Koller et al., 1987; Zoeller et al., 1993). The functional relationships among levels of this endocrine axis are so tightly linked that perturbations within one level produce compensatory changes in the other levels. The details of the relationship within this axis have not been explored fully for all non-mammalian vertebrates.

Thyroid Hormone Actions:

The majority of biological actions of thyroid hormones, including the regulation of brain development, are believed to be mediated by nuclear receptors for triiodothyronine (T3) (Lazar, 1993). Although the responsiveness to thyroid hormone requires the presence of nuclear TRs, the specific effects of thyroid hormone vary from tissue to tissue (Schwartz, 1983). Pleiotropic effects of thyroid hormone may be in part attributable to different levels and combinations of TR isoform expression (Lazar, 1993; Lazar, 1994). However, an important mechanism by which thyroid hormone effects can be regulated within cells, tissue, and across developmental stages is the interaction between receptors for thyroid hormone and those for retinoids (Forman and Samuels, 1990; Kliewer et al., 1992; Mano et al., 1994; Yu et al., 1991; Zhang et al., 1992). The implication of these observations is that thyroid hormone action can be modified, even disrupted, by agents which interfere with retinoid metabolism.

Despite the recognition that thyroid hormone exerts its effects through nuclear receptors, there are very clearly defined endpoints of thyroid hormone action during development. There are a few genes expressed in mammals whose expression has been rigorously defined as directly regulated by thyroid hormone in the mammal. These include myelin basic protein (MBP) (Mitsubashi et al., 1988) neuroganin/RC3 (Iniguez et al., 1993), TRH (Hollenberg et al., 1995), malic enzyme (Song et al., 1986), thyrotropin (Carr et al., 1993), and some neuron-specific genes (Thompson, 1996). In amphibians, a number of genes have been identified in frogs (*Xenopus*

laevis), which are shown to mediate effects of thyroid hormone on metamorphosis (Brown et al., 1996; Brown et al., 1995; Denver et al., 1997; Furlow et al., 1997; Kanamori and Brown, 1996), as well as more apical endpoints such as tail resorption.

Mechanisms of Antithyroid Activity:

A variety of environmental compounds are known to affect thyroid function or thyroid hormone action (Gaitan and Cooksey, 1989; Green, 1996). Processes known to be affected and a brief description of the effects are included below:

a) *Active transport of iodide into the thyroid gland*. Inhibitors include complex anions (e.g., ClO4, TcO4, thiocyanate). Reduces thyroid iodide uptake and can reduce thyroid hormone synthesis and circulating levels. Elevated TSH can overcome modest inhibition in the absence of other thyroid pathologies.
b) *Iodination of thyroglobulin (by thyroid peroxidase (TPO))*. Inhibitors include thionamides (e.g., propylthiouracil, methiazole, carbimazole), thiocyanate, aniline derivatives such as sulfonamides, substituted phenols (resorcinol), flavonoids, and iodide. Reduces thyroid hormone synthesis and circulating levels, but can be overcome by elevated TSH.
c) *Coupling reaction*. Iodinated tyrosine residues of thyroglobulin must be coupled by an ether linkage to form iodothyronines, which are released from the thyroid gland. Inhibition of this coupling reaction reduces thyroid hormone synthesis. This reaction may be controlled by TPO itself. Inhibitors include thionamides and other inhibitors of iodination, minocycline, and lithium salts.
d) *Hormone release*. This is a cAMP-dependent process stimulated by TSH. Inhibitors include iodide and lithium salts.
e) *Iodotyrosine deiodination*. This process is important for recovery of iodide within the thyroid gland. Inhibition causes the reduction in thyroid iodide content and thus, inhibition of thyroid hormone synthesis. Inhibitors include nitrotyrosines.
f) *Iodothyronine deiodination*. This reaction is important for conversion of thyroxine to the hormonally active tri-iodothyronine, and for the conversion of T3 to the hormonally inactive T2. Inhibitors include thiouracil derivatives, oral cholecystographic agents, and amiodarone.
g) *Hormone excretion or inactivation*. This process is affected by inducers of hepatic drug-metabolizing enzymes. Inhibitors include phenobarbital, phenytoin, carbamazepine, rifampicin, and organochlorines.
h) *Hormone action*. Thyroid hormone action is largely mediated by binding to specific nuclear receptors. There is limited evidence that compounds such as phenytoin (dilantin) and amiodarone can displace T3 from nuclear binding sites *in vitro*. However, there is little *in vivo* evidence that this interaction may compromise T3 action. In addition, there are predictions that specific PCBs may interfere with T3 binding to the nuclear receptor because of similarities in structure. However, these predictions have not yet been experimentally verified.

Recommended T1S Assays for Anti-Thyroid Activity:

To the EDSTAC's knowledge, all known antithyroid compounds so far reported in vertebrates affect circulating levels of thyroxine (T4). The physiological consequences of these effects are variable and may require considerable time to develop in a screening paradigm. In addition, they may represent endpoint measures that are not solely responsive to thyroid disruption. Therefore, the recommended T1S battery includes *in vivo* measures of circulating levels of thyroxine and TSH, and changes in the histopathology of the thyroid gland, as initial endpoints for antithyroid screening. These measures can be made in animals that are treated in the commission of other *in vivo* screens (e.g., uterotrophic assay). Human serum-based commercial kits are available to measure T4 and TSH by radio-immunoassay, and these assays have been validated and standardized for some other species. It is important to recognize that a significant change in circulating thyroxine or TSH should be considered a positive finding. This determination is based upon the fact that some compounds, such as PCBs, have been reported to reduce circulating levels of T4, but leave TSH unaffected (Goldey et al., 1995). In addition, weak antithyroid agents, especially those affecting some aspects of iodide metabolism in the thyroid (Gaitan and Cooksey, 1989; Gaitan et al., 1989; Green, 1996), may be compensated for by elevated TSH. Thus, T4 may appear normal, but TSH would be elevated. Finally, the absence of an effect on circulating levels of T4 or TSH does not preclude the possibility that an agent is antithyroid. It is well known that goitrogens can affect thyroid function over long periods and not be manifested by significant changes in circulating levels of T4 or TSH measured by radio-immunoassay (Gaitan et al., 1989). These compounds would produce a measurable effect on the thyroid gland. For these reasons, the EDSTAC recommends thyroid histopathologic evaluations.

During their deliberations, the STWG extensively discussed the timing of exposure to a chemical substance or mixture. EDSTAC recommends evaluation of antithyroid effects in animals prepared for testing other actions (either 14-day or 20-day exposure). Although no cases are known in which exposure to xenobiotics of greater than 14 days are required to significantly affect circulating levels of T4, TSH, or thyroid histopathology, EDSTAC believes longer periods may be required (DeVito et al., 1998). The effects of duration of chemical exposure must be quickly evaluated in the validation phase.

These measures in mammals represent evaluation of thyroid function; there are no clear markers of thyroid hormone action that could be used within the context of a T1S assay. In contrast, tail resorption in amphibian metamorphosis represents an assay which utilizes specific thyroid hormone-dependent effects as an endpoint for a T1S assay.

2. *In Vivo* Assays Using Other Vertebrates

The T1S battery includes an amphibian and a fish assay, which fill important needs in the battery and complements the information from assays using mammals. These assays help the battery meet design criteria 2 and 4, which express the need for a sufficient range of taxonomic subjects and range of metabolic functions be evaluated in the battery. While the basic biochemical processes of receptor binding and cellular activation by hormones are known to be similar among many organisms, detailed comparative data do not exist to assess the extent of the homology across vertebrate classes. In particular, fish ER differs from mammalian ER more than the ER of other classes, and fish have some unique androgens. Hence, including fish as subjects makes sense as it

is the class most likely to show differences from mammals in EAT activity. In addition, there are known differences in the ability of organisms to metabolize xenobiotics, due partially to the route of exposure. Fish and amphibians receive more dermal exposure to chemical substances or mixtures than other vertebrate classes, and thus chemical substances or mixtures avoid immediate metabolism in the liver. These assays also meet the need for the battery to have a clear cut response to measure the effects of thyroid hormone, which frog metamorphosis does, and hence complements the data on serum concentration of thyroid hormones and thyroid gland histology derived from the pubertal female rodent assay.

Unlike the mammalian assays, the assays recommended for fish and frogs have not been used in regulatory testing, and hence they need more work before being implemented for that purpose. In fact, however, both procedures using these species and similar endpoints have been used to investigate endocrine activity of chemical substances or mixtures in the research literature. The work needed for standardization of protocols and validation of the assay with known endocrine disruptors should proceed as soon as possible as these assays play a crucial role in the T1S battery. In addition, EDSTAC encourages development of other assays in the event that either of these two fail to be adequately standardized and validated, so that a complete screening battery can be implemented.

E. *In Vivo* Assay Overviews

Several measures of estrogenicity (reviewed by Gray et al., 1997a; Reel et al., 1996; Parker, 1966, Chapter 30) have been used for over 70 years, including uterine size, vaginal cornification, female sexual receptivity, and age at puberty/vaginal opening (see Parker, 1966, for a thorough review). For example, Dodds et al., 1938, found that DES produced full estrus in ovariectomized rats, so far as vaginal, uterine, and mating reactions were concerned. These remain some of the most useful short-term *in vivo* methods for screening for estrogenicity. Studies of xenoestrogens typically indicate that the sensitivity of these endpoints is as follows: uterine weight measured 5 hours after the last treatment, with fluid, is generally more sensitive than the age at vaginal opening or vaginal cornification; however, this is not always the case. Uterine histology and biochemical measures appear to be at least as sensitive to estrogens as uterine weight, but these endpoints are slightly more difficult to evaluate as they require specialized skills and equipment and are more expensive.

The sensitivity of the age at vaginal opening to methoxychlor appears to be about twofold greater than the onset of persistent vaginal cornification (PVC) and at least equivalent to the sensitivity of the uterotrophic assay (Gray et al., 1989; Gray and Ostby, 1998). However, in another study hydroxylated PCBs induced vaginal cornification at dosage levels that failed to induce an increase in uterine weight (Gillesby and Zacharewski, 1996). PVC was not detected in a long term study of the estrogenicity of octylphenol that doubled uterine weight in long-term ovariectomized rats after 10 weeks of oral administration and after three days of administration in juvenile rats (Gray and Ostby, 1998). Hence, some of the original measures of estrogenicity, in use now for nearly three-quarters of a century, are still regarded as the most useful indicators of estrogenic activity *in vivo*.

Recent studies to evaluate methoxychlor, 4-tert-octylphenol, nonylphenol, bisphenol A, DES (Reel et al., 1985), estrogens, and antiestrogens (Conner et al., 1996) have demonstrated the utility of these biological assays, since: (1) pro-estrogens/metabolites may be detected following *in vivo* exposure; (2) agonistic/antagonistic properties may be addressed; (3) bioaccumulation and/or the development of tolerance to exposure may be evaluated; (4) multiple routes and lengths of exposure may be easily compared; and (5) acute exposure regimes may be used. However, care should be taken when interpreting results from these biological assays since: (1) some environmental chemicals do not test positive for all measures; (2) exposure route and time of assessment following exposure affect the results; and (3) the observed biological response may result from other mechanisms of action. For example, in the ovariectomized female, increased uterine weight can be induced by aromatizable and nonaromatizable androgens (Salamon, 1938) and EGF (Nelson et al., 1991).

In the intact (ovaries present versus castrate or ovariectomized) juvenile female rat, the age at which treatment is initiated (typically 19-21 days of age) and the duration of treatment are critical variables that affect uterine weight. Exposure duration longer than 3 or 4 days or the use of juvenile females 24-25 days of age at the start of the study are not recommended because of the potential confounding of the treatment effect with the onset of natural estrous cyclicity and its concurrent fluctuations in uterine weight and histology. As long as the uterine weight bioassay has been used, it still has not been completely standardized, a fact that leads to some variation in results from lab to lab. For example, there are differences with respect to how thoroughly the mesenteric fat along the uterine curvature is removed, and some labs weigh the uterus with its contents, while others remove the fluid before weighing. Uterine weight, serum hormone concentrations, and other evaluations in intact female rats are difficult to interpret unless great care is taken to assure that females are necropsied at the same stage of the estrous cycle. With regard to the measurement of serum hormones in the cycling female rat, the time of day is also critical, in addition to the day of the cycle. Effects on estrous cyclicity are not limited to ER-mediated alterations; several other reproductive (hypothalamic-pituitary) and nonreproductive (hypothyroidism) endocrine-related alterations can alter estrous cyclicity in the female rodent. The detection of vaginal cornification in juvenile, and ovariectomized adult rodents is one of the original assays used to detect estrogenicity and, as indicated above, this assay appears to be relatively sensitive to weak estrogens. However, higher levels of xenoestrogens are required to disrupt estrous cyclicity and induce constant estrus PVC in intact female rats (Gray et al., 1989).

1. Rodent 3-Day Uterotrophic Assay (Subcutaneous)

Assay for estrogenicity

An increase in uterine weight is generally considered to be one of the best indicators of estrogenicity when measured in the ovariectomized (ovx) or immature female rat or mouse after 1-3 days of treatment. The recommended 3-day uterotrophic assay (sc injection) uses the ovariectomized adult female rat (the duration can be extended if so desired) with n =10/group. Subcutaneous treatment is recommended at this time because most of the historical data are collected in this manner and there is relatively few data concerning the effects of other routes of administration at this time. At necropsy one should carefully trim the uterus of fat and weigh with and without fluid and save uterus and vaginal tissues for histopathology. Most xenoestrogens have been examined in this assay. It also should be

executed in a manner to detect antiestrogens. In this regard, a control and xenobiotic-treated group are coadministered with estradiol sc and necropsied.

2. Rodent 20-Day Pubertal Female Assay With Thyroid

Assay for thyroid, HPG axis, aromatase, and estrogens that are only effective orally or after longer dosing than the uterotrophic assay.

The determination of the ages at "puberty" in the female rat is an endpoint that has already gained acceptance in the toxicology community. Vaginal opening (VO) in the female is a required endpoint measured in the new EPA two-generation reproductive toxicity test guideline. In this regard, this assay would be easy to implement because these endpoints have been standardized and validated and VO data are currently being collected under GLP conditions in most toxicology laboratories. In addition, VO data are reported in many recently published developmental and reproductive toxicity studies (i.e., see studies from R.E. Peterson's, R. Chapin's and L.E. Gray's laboratories on dioxins, PCBs, antiandrogens, and xenoestrogens).

In the pubertal female assay, oral dosing is initiated in weanling rats at 21 days of age (10 per group, selected for uniform body weights at weaning to reduce variance). The animals are dosed daily, 7 days a week, and examined daily for vaginal opening (one could also check for age at first estrus and onset of estrous cyclicity). Dosing continues until VO is attained in all females (typically two weeks after weaning, unless delayed). Age at VO is also determined in the female rat. Rats are dosed by gavage with xenobiotic and examined daily for VO. The advantage over the uterotrophic assay is that one test detects both agonists and antagonists, it detects xenoestrogens like methoxychlor that are almost inactive via sc injection, it detects aromatase inhibitors, altered HPG function, and unusual chemicals like betasitosterol. In addition, at necropsy one should weigh the ovary (increased in size with aromatase inhibitors, but reduced with betasitosterol), save the thyroid for histopathology, take serum for T4, and measure TSH.

Exposure of weanling female rats to environmental estrogens can result in alterations of pubertal development (Ramirez and Sawyer, 1964). Exposure to a weakly estrogenic pesticide after weaning and through puberty induces pseudoprecocious puberty (accelerated vaginal opening without an effect on the onset of estrous cyclicity) after only a few days of exposure (Gray et al., 1989). Pubertal alterations also result in girls exposed to estrogen-containing creams or drugs, which induce pseudoprecocious puberty and alterations of bone development (Hannon et al., 1987).

Several examples of estrogenic chemicals affecting vaginal opening in rodents are known and include methoxychlor (Gray et al., 1989), nonylphenol, and octylphenol (Gray and Ostby, 1998). This endpoint appears to be almost as sensitive as the uterine weight bioassay, but the evaluation is easier to conduct and does not require that the animals be euthanized, so they can be used for additional evaluations. For example, treatment with methoxychlor at weaning (6 mg/kg/day or higher) caused pseudoprecocious puberty in female rats. Vaginal opening occurs from two to seven days earlier in treated animals than controls, in a dose-related fashion, but methoxychlor did not alter estrous cyclicity at the low dosage levels, indicating a direct estrogenic effect of methoxychlor on vaginal epithelial cell function without an effect on hypothalamic-pituitary maturation. Similar effects have been achieved

with chlordecone, another weakly estrogenic pesticide, and octylphenol. Chlordecone also induces neurotoxic effects (hyperactivity to handling and tremors). In addition to estrogens, the age at vaginal opening and uterine growth can be affected by alteration of several other endocrine mechanisms, including alterations of the hypothalamic-pituitary-gonadal axis (Shaban and Terranova, 1986; and Gonzalez et al., 1983). In rats, this event can also be induced by androgens (Salamon, 1938) and EGF (Nelson et al., 1991). In the last 20 years there have been over 200 publications which demonstrate the broad utility of this assay to identify altered estrogen synthesis, ER action, growth hormone, prolactin, FSH or LH secretion, or CNS lesions.

3. Rodent 5-7 Day Hershberger Assay

Assay for Antiandrogens

In the castrated male rat, the gonads have been removed and effects on androgen-dependent accessory sex organs and tissues are likely to be direct and not a result of pituitary or gonadal secretion. The assay (Hershberger, 1953) requires two stages as below:

- castrated male rat + T + Xenobiotic (to detect antagonist)
- castrated male + X (to detect agonist)

In this *in vivo* test, sex accessory gland weights (ventral prostate and seminal vesicle separately) are measured in castrated, testosterone-treated adult male rats after 4-7 days of treatment by gavage with the test compound. The advantage of this assay is that it is fairly simple, short term, and relatively specific compared to other *in vivo* procedures. Although the androgens, testosterone, and dihydrotestosterone (DHT), play a predominant role in the growth and maintenance of the size of these structures, several other hormones and growth factors can influence sex organ weights including the thyroid and growth hormones, prolactin, and epidermal growth factor (EGF) (Luke and Coffey, 1994). Exposure to estrogenic pesticides can also reduce sex accessory gland size; however, it is unclear to what degree these reductions result from direct versus indirect action of the chemical. Other useful endpoints that help reveal the mechanism of action include serum hormone levels of T, DHT, LH, AR distribution, TRPM2/C3 gene activation, ODC, and 5-alpha-reductase activity in the prostate. The prostate and seminal vesicles should be weighed separately because these organs differ with respect to the androgen that controls their growth and differentiation. The prostate is dependent upon enzymatic activation of T to DHT, whereas the seminal vesicle is less dependent upon this conversion. Hence, effects on 5-alpha-reductase can be distinguished from AR-mediated mechanisms by determining whether the prostate is preferentially affected. Growth of the levator ani muscle is T dependent, having little capacity to convert T to the more potent androgen DHT. Weight of this muscle is useful in identifying anabolic androgens and antiandrogens, and for this reason has been used extensively in the pharmaceutical industry. In order to detect androgenic rather than antiandrogenic action one would simply delete the hormone administration from the protocol.

Data from this assay (often with slight modifications), using drugs and xenoantiandrogens, are widely available in the literature. For a non-*in utero* assay, this assay robustly detects androgens and antiandrogens with a dynamic response that typically exceeds that of the intact adult male (Raynaud, 1984). Most of the studies are able to detect significant effects with only five animals per group. In

fact, in one study which used 10-15 drugs, the Hershberger assay was more responsive than was the intact male for every chemical examined (Wakeling et al., 1981). The power of the castrate-male assay arises from the fact that castration creates a "fetal-like" endocrine system with regard to the regulation of androgen secretion, as the HPG axis can no longer compensate for the effect of the chemical on the AR. For example, p,p' DDE reduces sex accessory gland weights in this assay, but not when administered to intact male rats (Kelce et al., 1997).

4. Frog Metamorphosis Assay

This assay employs intact larval (tadpole) stages of the African clawed frog *(Xenopus laevis)* exposed over a 14-day time period, 50-64 days of age, to observe the rate of tail resorption (Fort and Stover, 1997). Tail resorption can be easily quantified with computer-aided video image processing (Fort and Stover, 1997). The molecular mechanisms involved in tail resorption are well characterized (Brown et al., 1995; Hayes, 1997a) and this assay is, therefore, considered to be a simple and specific assay for thyroid action. It will detect thyroid (increase in tail resorption rate) and antithyroid (decrease in tail resorption rate) effects. Because evidence also suggests that thyroid action on tail resorption is regulated by corticoids, estrogens, and prolactin (Hayes, 1997b), this assay will address distinctive modulating pathways and, in tandem with the 14-day mammalian pubertal assay, a comprehensive screen for thyroid hormone activity is achieved.

5. Fish Gonadal Recrudescence Assay

Intact mature fish maintained under simulated "winter" conditions (short day length, cool temperatures) exhibit regressed secondary sex characteristics and gonad maturation. In this assay, intact fish of both sexes (fathead minnow, *Pimephales promelas,* or other appropriate species) are simultaneously subjected to an increasing photoperiod/temperature regime and test substance to determine potential effects on maturation from the regressed position (recrudescence). The primary endpoints examined in the assay include morphological development of secondary sexual characteristics, ovary and testis development (weight increases), gonadosomatic index (ratio of gonadal weight to body weight), final gamete maturation (ovulation, spermiation), and induction of vitellogenin. This assay is sensitive to HPG axis effects in addition to androgen- and estrogen-related activity.

Fish differ in steroid profiles from mammals (e.g., 11-ketotestosterone as opposed to testosterone is the most important androgen in fish). The estrogen receptor in fish appears to differ structurally and functionally from the mammalian estrogen receptor (Petit et al., 1995; Gustafsson, 1996). Also, steroid receptors in eggs and hepatic vitellogenin production have no known analogous receptors in mammals, which would suggest sites of endocrine disruption unique to oviparous animals. Therefore, this assay is essential to address these known endocrine differences.

F. Alternative Assays for Possible Inclusion

1. Placental Aromatase Assay

One critical enzyme present at very low levels in the testis, and at higher levels in the ovary, uterus, and placenta, is aromatase, which converts testosterone to estradiol and is another P450 isozyme. Human placental aromatase is commercially available and could be used *in vitro* to assess the effects of toxicants on this enzyme fairly easily.

2. Modified Rodent 3-Day Uterotrophic Assay (Intraperitoneal)

This is an *in vivo* assay (O'Conner et al., 1996) for estrogenic activity in ovariectomized female rats. It can detect certain antiestrogens with mixed activity, i.e., those with some agonistic activity (e.g., tamoxifen). The rats are injected intraperitoneally with the test agent daily for three days. The ip injection method may enhance the sensitivity of the assay and is capable of detecting the estrogenic potential of methoxychlor, which has been cited as an example of a compound not detectable by the sc route. The females are necropsied either 6 hours or 24 hours after the final treatment, depending on the protocol employed by the laboratory. Vaginal cytology is evaluated by vaginal lavage to determine whether the epithelium has become cornified, indicative of estrus. Presence of fluid in the uterine lumen is noted and recorded, and the number of animals that have fluid in the uterus is reported. Fluid imbibition (uptake) is indicative of estrogenic potential. The uterus is excised and weighed. It is then preserved in an appropriate fixative for subsequent histological evaluation, if needed.

Subsequent histological evaluation will be triggered by an equivocal uterine weight or uterine fluid response (i.e., an increase that is not statistically significant). This evaluation will consist of a characterization of the appearance of the uterine epithelium, a measurement of uterine epithelial cell height, and epithelial mitotic index or proliferating cell nuclear antigen (PCNA) immunohistochemistry. Uterine cell height and cell proliferation are sensitive indicators of estrogenic potential.

3. 14-Day Intact Adult Male Assay

This *in vivo* assay is intended to detect effects on male reproductive organs that are sensitive to antiandrogens and agents that inhibit testosterone synthesis or inhibit 5-alpha-reductase (Cook et al., 1997). The duration of the assay is anticipated to be sufficient to detect effects on thyroid gland activity. The rats are anatomically intact and mature; therefore, they have an intact HPG axis, allowing an assessment of the higher order neuroendocrine control of male reproductive function and the thyroid.

Young adult male rats (70-90 days of age) are used in this assay. They are dosed daily with the test agent for 14 days. The recommended route of administration is ip, which may, in some cases, maximize the sensitivity of the assay. They are necropsied 24 hours after the final dose. Immediately after sacrifice one cauda epididymis is weighed and processed for evaluation of sperm motility and concentration. The following organs are weighed: testes, epididymides,

seminal vesicles, and prostate. The following are fixed and evaluated histologically: one testis and epididymis, and the thyroid. The following hormones are measured in blood plasma: T4, TSH, LH, testosterone, DHT, and estradiol.

Empirical assessment of this assay has shown it to be sensitive to agents that are directly antiandrogenic, inhibit 5-alpha-reductase, inhibit testosterone synthesis, or affect thyroid function. The sensitivity of this assay, as defined as the ability to detect a hazard, may be comparable to other assays that have been recommended.

4. Rodent 20-Day Thyroid/Pubertal Male Assay

This assay detects androgens and antiandrogens *in vivo* in a single stage apical test. "Puberty" is measured in male rats by determining age at PPS (preputial separation). Animals are dosed by gavage beginning one week before puberty (which occurs at about 40 days of age) and PPS is measured. Androgens will accelerate and antiandrogens and estrogens will delay PPS. The assay takes about 3 weeks, and allows for comprehensive assessment of the entire endocrine system in one study (10 per group, selected for uniform body weights to reduce variance). The animals are dosed daily, seven days a week, and examined daily for PPS. Dosing continues until 53 days of age; the males are then necropsied. The body, heart (thyroid), adrenal, testis, seminal vesicle plus coagulating glands (with fluid), ventral prostate, and levator ani plus bulbocavernosus muscles (as a unit) are weighed. The thyroid is retained for histopathology and serum is taken for T4, T3, and TSH. Testosterone, LH, prolactin, and dihydrotestosterone analyses are optional. These endpoints take several weeks to evaluate and are affected not only by estrogens but by environmental antiandrogens, drugs that affect the hypothalamic-pituitary axis (Hostetter and Piacsek, 1977; Ramaley and Phares, 1983), and by prenatal exposure to TCDD (Gray et al., 1995a; Bjerke and Peterson, 1994) or dioxin-like PCBs (Gray et al., 1995b). In contrast to these other mechanisms, only peripubertal estrogen administration accelerates this process in the female and delays it in the male. Preputial separation in the male rodent is easy to measure and this is not a terminal measure (Korenbrot et al., 1977).

Age and weight at puberty, reproductive organ weights, and serum hormone levels can also be measured. Delays in male puberty result from exposure to both estrogenic and antiandrogenic chemicals including methoxychlor (Gray et al., 1989), vinclozolin (Anderson et al., 1995b) and p,p' DDE (Kelce et al., 1995). Exposing weanling male rats to the antiandrogenic pesticides p,p' DDE or vinclozolin delays pubertal development in weanling male rats as indicated by delayed preputial separation and increased body weight (because they are older and larger) at puberty. In contrast to the delays associated with exposure to estrogenic substances, antiandrogens do not inhibit food consumption or retard growth (Anderson et al., 1995). Antiandrogens cause a delay in preputial separation and affect a number of endocrine and morphological parameters including reduced seminal vesicle, ventral prostate, and epididymal weights. It is apparent that PPS is more sensitive than are organ weights in this assays. In addition, responses of the HPG are variable. In studies of vinclozolin, increases in serum LH were a sensitive response to this antiandrogen, whereas serum LH is not increased in males exposed to p,p' DDE during puberty (Kelce et al., 1997). Furthermore, a systematic review of the literature indicates that the sex accessory glands of the immature intact male rat are consistently more affected than in the adult intact male rat.

In summary, preputial separation and sex accessory gland weights are sensitive endpoints. However, a delay in preputial separation is not pathognomonic for antiandrogens. Pubertal alterations result from chemicals that disrupt hypothalamic-pituitary function (Huhtaniemi et al., 1986), and, for this reason, additional *in vivo* and *in vitro* tests are needed to identify the mechanism of action responsible for the pubertal alterations. For example, alterations of prolactin, growth hormone, gonadotrophin (LH and FSH) secretion, or hypothalamic lesions alter the rate of pubertal maturation in weanling rats.

As indicated above, the determination of the age at "puberty" in the male rat are endpoints that already have gained acceptance in the toxicology community. Preputial separation in the male is a required endpoint in the new EPA two-generation reproductive toxicity test guideline. In this regard, this assay would be easy to implement because these endpoints have been standardized and validated and PPS data are currently being collected under GLP conditions in most toxicology laboratories. In addition, PPS data are reported in many recently published developmental and reproductive toxicity studies (i.e., see studies from R.E. Peterson's, J. Ashby's, R. Chapin's and L.E. Gray's laboratories on dioxins, PCBs, antiandrogens, and xenoestrogens).

Sex accessory gland weights in intact adult male rats also can be affected directly or indirectly by toxicant exposure. The HPG axis in an intact animal is able to compensate for the action of antiandrogens by increasing hormone production, which counteracts the effect of the antiandrogen on the tract (Raynoud et al., 1984; Edgren, 1994; Hershberger, 1953).

IV. General Principles in Evaluating Tier 1 and Tier 2 Results

A. Introduction

Apart from substances yielding negative results in all assays, it is likely that most substances will produce a unique array of results requiring a judgment as to whether the weight of evidence indicates the substance should or should not be judged a candidate for T2T (after completing T1S), and, designated as an endocrine disruptor for EAT (after completing T2T). A table consisting of 18 chemical types along with known or expected T1S results can be found in Appendix P, Examples of "Weight-of-Evidence" Determinations.

There are two senses in which a "weight-of-evidence" determination will need to be made. The first is with respect to the question of whether consistent results are being obtained across multiple assays. If the results are not consistent, it will be necessary to "weight" the conflicting results, allowing some to carry more weight than others. The second sense is with respect to the question of whether a particular body of evidence, even if it is fully consistent, is sufficient to justify a decision. In this sense, it is the "weight" of the entire body of evidence, relative to some minimal level established as being required for sound decisions, that is being judged.

Assessing the "weight-of-evidence," and using that assessment in forming judgments about a substance, can be done in a variety of ways. On one extreme are approaches based solely on expert judgment in which an individual reflects on the data and offers an informed, yet personal,

opinion. On the other extreme are more formal and mathematical procedures such as Bayesian analysis in which data are viewed sequentially and used to formulate a priori and a posteriori judgments. An intermediate approach is one in which a group debates the available data, presents alternative arguments, and collectively reaches a judgment.

All three of these possibilities are forms of "weight-of-evidence" assessments. The EDSTAC has agreed not to prescribe a particular "weight-of-evidence" approach, as these are controversial and a matter of science policy to be established by the Agency. Instead, the EDSTAC offers general guidelines for reasoning from the data produced in the two tiers, which conform to the outline provided in the NAS/NRC report Science and Judgment in Risk Assessment (National Academy Press, 1994). These guidelines provide a framework within which one may take into account multiple features of relevant screening and testing data in determining whether the substance should be a priority for T2T (after T1S) and/or is determined to have endocrine disrupting effects (after T2T).

"Weight-of-evidence" considerations will arise at two places within the EDSTP. It first will arise in considering whether the evidence collected solely within a given tier (T1S) warrants a particular conclusion (e.g., that the substance may have endocrine activity for EAT). The second place where it will arise is in considering whether results from the previous tier (T1S) should affect the conclusions drawn from the subsequent tier (T2T). By this, the EDSTAC is not referring to the fact that T1S results may guide selection and/or design of Tier 2 tests (with the results of the Tier 2 tests then being interpreted without further reference to the T1S results). The EDSTAC is, instead, referring to the possibility that the T1S assay results may be "weighted into" the determination of whether a substance has passed or failed the Tier 2 tests.

A broad range of results may need to be weighted into a final judgment at either tier. Information routinely taken into consideration in determining the "weight-of-evidence" will include:

- the balance of assays/tests that gave positive and negative results;
- results of *in vitro* versus *in vivo* assays/tests;
- the nature of the biological effects induced;
- the range of effects observed;
- the slope and shape of the dose-response curves;
- the level, magnitude, or severity of the effects induced; and
- the presence or absence of response in multiple taxa.

The "weight-of-evidence" approach makes explicit the assumption that results of some assays/tests, in some taxa, at some level of severity, are intrinsically "worth" more than others and should, therefore, carry more weight in decisions following T1S and T2T. For example, positive results showing reproducible, high levels of effects at low doses (near the doses produced by environmental or human exposures) are likely of greater weight than weak effects observed only at very high, perhaps excessively toxic, levels of exposure.

The EDSTAC has taken the approach here of providing guidance on the use of "weight-of-evidence." Any approach used must satisfy several broad criteria which are essential. The

weighting system should be transparent, allowing individuals to review the "weight-of-evidence" determination. It should be possible to understand the procedure before viewing the data, so individuals have a reasonable expectation of the final decision at the time when the data are presented. This does not mean the decision is fully determined by the data, removing the need for scientific judgment, but it does mean that any deviations from the expected decision should be supported by an explanation detailing the "weight-of-evidence" assigned.

B. False Negatives and False Positives Within the Context of T1S and T2T

The guiding principle for the treatment of false positives and negatives should be one of valuing sensitivity more than specificity at the screening level unless this compromises the ability to sort chemical substances or mixtures into a subset most likely to be of concern. False positives and negatives can arise in at least three different ways in the screening and testing batteries (see Figure 5.1):

- The false result may be due to the stochastic nature of screens and tests. A false result leading to an incorrect claim that the screen/test is positive is a Type I (false positive) error. A false result leading to an incorrect claim that the screen/test is negative is a Type II (false negative) error. The frequency of these types of errors is expressed by the p value[2] for an assay, so the selection of a required p value to classify a result as positive will determine the frequency of Type I and Type II errors. The guiding principle above suggests that required p values should be chosen so Type II errors are minimized, while also ensuring that Type I errors do not become so frequent that chemical substances or mixtures can no longer be sorted meaningfully.

- False positives and negatives may arise due to unknown or unexpected limitations of the test or assays, such as anomalous activity of chemicals or classes in a particular assay or interference from assay procedures.

- The third source of error arises from a potential lack of predictivity of results in T1S for endocrine disruptive responses in T2T. This source of error is shown in Figure 5.1 by the bold arrow going from positive results in T1S to T2T. A negative result in T1S may simply mean the assay battery misses a mechanism of action that would have been active in a Tier 2 Test. This will result in a false classification of the substance as not having endocrine activity, an error that would have been caught in T2T had the chemical substance or mixture proceeded to that stage. For this reason, the T1S battery was designed to capture all known endocrine mechanisms for EAT and to minimize false negative results specifically as opposed to false positives. A positive result in T1S could be followed by negative results in T2T because the endpoints measured in T1S may not accurately predict adverse effects in long-term, whole animal tests. This will result in unnecessary testing of some chemicals in T2T, a possibility considered more acceptable than missing potential endocrine disruptors for EAT.

In treating the frequency of Type I and Type II errors, it is important to consider both the frequency of these errors in each particular assay/test and the number of assays/tests in a battery. As the number of assays/tests in a battery increases, the probability that *at least one* of the assays/tests will show a false positive increases. This is shown in Figure 5.2, which displays the

[2] "P value" is the statistical probability that two groups, e.g., control and treated animals, come from the same population (i.e., p=0.05 means that the probability is 5 out of 100 that the two groups, based on values for a given parameter, belong to the same population and conversely, that 95 out of 100 do not belong to the same population and therefore they are significantly different).

relationship between the probability of any one assay/test showing a false positive (the X axis), the number of assays/tests in the battery (the Y axis), and the probability that the battery shows at least one false positive result. In this figure, it is assumed that the chemical substance or mixture tested actually has no endocrine activity, but might yield a false positive result due primarily to stochastic variation. The goal of the Tier 1 or Tier 2 stages should be to minimize the probability of a battery producing a Type II error for a chemical substance or mixture, while not causing the probability of a Type I error from getting so large that the battery becomes ineffective at sorting chemical substances or mixtures. Figure 5.3 displays the analogue of Figure 5.2 (i.e., the probability that a battery produces a false negative result if each assay/test has a given false negative frequency, there are N assays/tests in the battery, and the chemical substance or mixture truly is an endocrine disruptor for EAT).

The EDSTAC cautions that the statistical properties of actual assays/tests in a battery will not be identical, so Figures 5.2 and 5.3 are simply illustrative and must be modified for any particular battery developed. What the figures indicate is that weighting a single positive result from a battery into the "weight-of-evidence" judgment should reflect a concern for both Type I and Type II errors. From these figures, it can be seen that a large battery (e.g., with 10 assays/tests), each with a false positive frequency of only 10%, can result in a very high probability of producing at least one assay/test showing a false positive when applied to a substance that in reality has no endocrine-disrupting properties. Such a battery would be essentially useless in sorting chemical substances or mixtures and focusing society's resources. The final advice here is that an effort should be made to characterize statistically the frequency of Type I and Type II errors associated with any selected battery, and to use this characterization in deciding the weight assigned to a single positive result from that battery.

C. Specific Principles for Evaluating T1S

There are several specific criteria to be met by the decision process assuming appropriate dose and route of exposure as discussed previously in this chapter:

1. If functionally equivalent information is available (e.g., from the sorting and prioritization phase), it may be appropriate that only those T1S assays which evaluate the endocrine activity of concern (based on prior information) of a chemical substance or mixture would be performed (i.e., only a subset of assays would be run). Similarly, the results of the T1S assays may require that only a subset of the Tier 2 tests be conducted.

2. If all assays are performed, and all assays are negative, then the chemical substance or mixture does not have endocrine activity for estrogen, androgen, or thyroid hormone at this time.

3. *In vitro* assays cannot and will not be "gatekeepers;" they cannot constitute a "decision node;" they are useful as information for possible mechanisms (or site of action) but not as "yes/no" determinants to proceed to the *in vivo* screens or T2T because:
   a) *in vitro* assays mediated by receptor binding evaluate only <u>one</u> of <u>many</u> possible sites and modes of action;
   b) negative results may mean relatively little due to limitations of the assays, e.g., lack of metabolic capability, solubility, etc. (i.e., false negatives); and
   c) positive results may be false positives.

4. Results from *in vivo* assays have more weight than results from *in vitro* assays since:
   a) *in vitro* assays will generate false negatives as well as false positives, based on differences in access to the target tissue, metabolism, etc. relative to *in vivo* assays; and
   b) *in vivo* results are considered to be more relevant.

5. Results from *in vitro* assays that assess endocrine activity with and without metabolic activation are worth more than results from *in vitro* assays without metabolic activation (since the former can assess the activity of metabolites generated within the culture if the correct metabolic activation is used (e.g., rat liver S9) and the latter can only assay the parent compound).

6. Results from apical *in vivo* assays are worth more than results from specific *in vivo* assays (since they indirectly assay many more sites of action to get to the same endpoint; e.g., uterotrophic assay in ovariectomized adult females [specific assay; chemical substance or mixture must act at level of uterus] versus in intact immature females [apical assay; chemical substance or mixture can act at level of hypothalamus, pituitary, gonad, thyroid, and/or uterus]). A positive specific assay provides mechanistic information but other mechanisms of action may also be present and go undetected; a negative specific assay is less informative.

7. Biologically plausible results are worth more than biologically implausible results (obviously dependent on the state of current scientific knowledge).

8. Statistical significance is a useful tool, but must be interpreted within the context of biological significance. For example, an observed association which does not achieve statistical significance, but which is consistent with results from related assays suggesting a common mechanism of action, might be interpreted as biologically significant. This means the use of any particular criterion such as p equal to 0.05 should be carefully considered, and there may be no hard and fast rule for weighting by statistical significance.

9. A consistent pattern of positive (or negative) results in various related assays is worth more than a single isolated positive (or negative) result (e.g., positive results for binding to ER and transcriptional activation *in vitro* and positive results in an apical or specific uterotrophic assay *in vivo* are worth more than a positive result for receptor binding and transcriptional activation, but no uterotrophic response) (see additional comments in discussion of false negatives and false positives above).

10. The decision which will emerge from T1S is:
    a) The chemical substance or mixture does not require further testing for EAT activity at this time (the chemical substance or mixture goes to the "hold box"); or
    b) The chemical substance or mixture should be tested further for EAT activity at this time, and
        i) proceed to T2T; or
        ii) proceed to Hazard Assessment/Regulatory Action.

V. Tier 2 Testing Concepts and Design Parameters

A. Introduction to T2T

The purpose of T2T is to characterize the nature, likelihood, and dose-response relationship of endocrine disruption of estrogen, androgen, and thyroid in humans and wildlife. T2T is a complement to T1S and T2T results supersede T1S results. As already discussed, T1S is composed of a battery of *in vitro* and *in vivo* assays designed to detect whether a substance may have EAT activity. The *in vitro* screening assays are highly sensitive and quite selective for a particular mode of action. They are, however, quite far removed from the biological complexity of an intact animal and may give false positive readings because, for instance, not all substances which bind to a receptor will cause an adverse biological effect; false negative readings may also result from the *in vitro* receptor binding or transcriptional assays because not all endocrine disruptors act via a receptor. *In vivo* assays encompass the metabolic and response capability of a whole organism but focus on such a short time frame that the full effects of exposure to a chemical substance may not be identified and characterized. Since there is considerable biological conservation in the endocrine system, it is not necessary to screen in every major taxonomic group. Screens based on mammalian cell lines or intact animals will determine whether a chemical substance or mixture may interact with the endocrine system, and if so, the effects of the chemical substance or mixture must be characterized in longer-term studies in species representing a variety of taxa.

T2T is the final phase of the screening and testing program and is intended to provide more detailed information regarding endocrine disruption activity of a tested chemical substance or mixture. This tier should assess the concentrations which elicit endocrine disruption and the consequences of such disruption to inform hazard assessments. To fulfill this purpose, tests are longer-term studies designed to encompass critical life stages and processes, a broad range of doses, and administration by a relevant route of exposure, so a more comprehensive profile of biological consequences of chemical exposure can be identified and related to the dose or exposure which caused them. Effects associated with endocrine disruption may be latent and not manifested until later in life or may not appear until the reproductive period is reached. Tests for endocrine disruption will usually encompass two generations including effects on fertility and mating, embryonic development, sensitive neonatal growth and development, and transformation from the juvenile life stage to sexual maturity unless a rationale exists to limit the mammalian test to one generation.

The outcome of T2T is designed to be conclusive in relation to the outcome of T1S, and any other prior information, in the sense that a negative outcome in T2T will supersede a positive outcome in T1S. Furthermore, each full test in T2T has been designed to include those endpoints that will allow one to reach a definitive conclusion as to whether or not the tested chemical substance or mixture is or is not an endocrine disruptor for EAT in that species/taxa. Conducting all five tests in the T2T battery would provide a more comprehensive profile of the effects a chemical substance or mixture could induce via EAT endocrine disruption mode(s)/mechanism(s) of action than would be the case if only a subset of tests were performed. The EDSTAC recommends that the "default" action, in the absence of any prior information, be to perform all tests in the T2T battery with all endpoints. However, performance of the entire battery with multiple generations may not always be necessary, as is discussed below.

The next section of this report provides guidance on the selection of the taxa to be subjected to T2T, and the use of alternative tests, focusing particularly on alternative mammalian tests. Conducting Tier 2 tests on less than the five recommended taxa and/or the use of an alternative mammalian test will result in T2T outcomes that are, by definition and design, less comprehensive in comparison to the outcomes from T2T using the recommended two-generation mammalian test and/or the full set of recommended taxa. However, the EDSTAC believes the guidance set forth below will ensure that the information generated from tests conducted with less than the full set of recommended taxa and/or using alternative Tier 2 tests will still be sufficient, along with prior information, for EPA to make a decision about whether the chemical substance or mixture should be placed in the "hold box" or forwarded to hazard assessment.

While two-generation tests are designed to fully characterize potential effects of concern, there may be instances (depending on available prior information) when a less comprehensive study design would provide adequate information on which to make decisions. In addition, there may be reasons why all of the non-mammalian tests need not be conducted in the same time frame, or at all. Below is a discussion of how the decisions of which tests to perform should be informed.

In determining which mammalian and non-mammalian tests to include in T2T, the EDSTAC chose tests which would require the least modification from existing standardized methods in order to minimize the time necessary to standardize and validate those tests. Thus, the Tier 2 tests for mammals (rats), fish, birds, and invertebrates are based on existing test guidelines (the recommended study design for amphibians has never been utilized in a testing scheme). The existing test guidelines were considered adequate for evaluating the most obvious consequences of EAT-related endocrine disruption, but not adequate to pick up some of the more subtle or insidious effects. However, more modification of the non-mammalian guidelines was required than for their mammalian counterparts and the elements described in the existing non-mammalian guidelines, in general, have not been widely and routinely run, as have the elements described in the mammalian guidelines.

B. Outline of Recommended T2T Battery

The EDSTAC recommends that the T2T battery include a mammalian two-generation reproductive toxicity study, or a less comprehensive mammalian test in accordance with guidelines outlined below, and tests addressing four additional taxonomic groups, including birds, amphibians, fish, and invertebrates as follows.

Mammalian Tests

1. Two-Generation Mammalian Reproductive Toxicity Study; or
2. A Less Comprehensive Test:
    a) Alternative Mammalian Reproductive Test; or
    b) One-Generation Test.

Multigeneration Tests in Other Taxa

1. Avian Reproduction (with bobwhite quail and mallard)
2. Fish Life Cycle (fathead minnow)
3. Mysid Life Cycle (*Americamysis*)
4. Amphibian Development and Reproduction (*Xenopus*)

C. Guidance for Selecting Tier 2 Tests

The Conceptual Framework, found in Chapter Three, states that existing information on biological effects and exposure and the results of T1S should be used to inform decisions regarding the selection and design of Tier 2 tests. The EDSTAC believes that T1S information may be of use in determining whether to evaluate thyroid effects, for example, but may be of limited value in determining whether all five, or some subset of the recommended Tier 2 tests should be conducted since only a limited number of taxa are recommended for the *in vivo* assays and mammalian cell lines are used for *in vitro* assays in T1S. Therefore, the EDSTAC recommends that the choice of whether Tier 2 tests will be conducted on all five of the recommend taxa, or a more limited subset of the five taxa, should be based on the physico-chemical characteristics and environmental release and exposure information of the chemical substance or mixture to be tested, together with biological data from T1S.

This section addresses the need for guidance in the selection of Tier 2 tests, focusing upon: (1) the determination of which of the five taxonomic groups should be included in the Tier 2 testing of a specific chemical substance or mixture; (2) the circumstances under which it may be appropriate to perform an alternative test, with a particular focus on the selection of alternative mammalian tests; (3) the selection of endpoints; (4) the special case of chemicals that bypass T1S and go directly to T2T; and (5) the potential need for supplemental information to complete T2T.

1. Determining Which Taxa Should be Included in Tier 2 Testing for a Specific Chemical

In general, the number of studies underway at any one time is limited, in part, by laboratory capacity. In addition, the pace at which each study can be completed is controlled by the nature of the protocol for the study. Priority should be given to conducting tests for those species/taxa to which exposure is known or expected to occur. The EDSTAC recommends that in the absence of information, as described below, the default assumption should be that chemical substances or mixtures should be subjected to the full set of five recommended tests with all endpoints, once they are fully standardized and validated. Building on these overarching principle, the EDSTAC recommends that EPA should use the following guidelines in determining which taxa/tests should be included T2T for a specific chemical substance or mixture:

a) Where use, exposure, and release of a substance are well characterized, it may be possible to tailor T2T for particular exposure scenarios. Conversely, if they are poorly characterized, positive results in T1S would trigger the entire battery of Tier 2 tests unless other data clearly indicated that certain taxa would not be affected. As noted above, the EDSTAC recommends the default assumption, in the absence of information that certain taxa would not be affected, be to conduct all five of the recommended tests with all endpoints.

b) If sufficient exposure data are available, the owner of the chemical may conduct an exposure assessment to provide information that will help inform the selection of Tier 2 tests. Exposure assessments vary in scope and complexity, but, for this purpose, "sufficient data" at a minimum includes chemical identity, basic chemical properties (water solubility, octanol: water partition coefficient, vapor pressure, Henry's Law constant), rates of significant transformation processes such as biodegradation, and use and release profiles, including accidental releases. Measured values are preferable for chemical properties and environmental fate, but estimation methods are often satisfactory for supplying missing data. Ideally, the chemical use and release profile should provide information on the distribution of releases (if any) to air, water, soil, etc., and the amounts and frequency of such releases.

   For example, freshwater aquatic exposure modeling often can be conducted using models which incorporate environmental degradation processes. One such model, the Probablistic Dilution Model, yields the frequency of exceedance of an ecological concern level preset by the modeler, and is useful not only when releases are from known point sources but also when only the category of use (via Standard Industrialization Classification Codes) is known. At a higher level of complexity, even site-specific models such as EXAMS-II may be appropriate under some circumstances, for example, when there are only a few point sources and the site of release and downstream environments are well characterized.

   Soil fate models such as SESOIL and PRZM exist but also have extensive parameterization requirements. Short of this, rough estimates of mobility in soil and thus likelihood of reaching groundwater or surface waters that are in hydrological contact with ground water can be

made using certain screening-level tools routinely applied by EPA in, for example, PMN review.

For air releases numerous fate models such as ISCLT are available, and like the water models, they vary in complexity. They calculate concentrations of chemicals in air at assumed locations of human receptors (e.g., at certain distances from stack releases), but such exposure data may also be used as input for aquatic or terrestrial ecological exposure assessments.

Models for calculating environmental concentrations of chemicals released to estuaries are less well developed in general, but the model ESTUARY contains extensive hydrologic data for several major estuaries in the U.S. and is potentially useful. It was designed for high-volume consumer products that are widely used and dispersed in the environment (i.e., surfactants).

c) If a chemical is released to or can be predicted to reach streams, rivers, or lakes, a fish life cycle test with freshwater species and invertebrate life cycle test should be conducted. If release is to an estuary, marine species should be substituted for freshwater species in toxicity testing. If release is to both types of environment, freshwater species are preferred. In a broad sense, freshwater fish species are often more sensitive than saltwater fish species in laboratory toxicity tests. The greater sensitivity is due, in part, to greater bioavailability in freshwater which has fewer dissolved solids than brackish or saltwater. Moreover, currently freshwater species are more widely available and somewhat easier to accommodate in laboratory settings than saltwater species. Species selection, independent of taxa, should be performance-based, and as more appropriate species are developed for use and validated, the EDSTAC strongly encourages their use.

d) Pesticides with agricultural or other outdoor uses, and chemicals that would be expected to bioaccumulate and biomagnify through the food chain or that present a potential exposure to birds and aquatic organisms, should be tested in the appropriate tests. Alternatively, if the pesticide or other chemical is not likely to be used in a manner that results in exposure to birds and aquatic organisms, it does not need to be tested in either the avian or aquatic tests. For example, as per current regulatory requirements under FIFRA (40 CFR 158), this would be the case if a pesticide registrant is requesting the first use of a pesticide only for application to flowers grown only in greenhouses. Since a greenhouse is considered a confined-use situation, the studies to assess the impact on birds or aquatic organisms would not be required. If, at a later time, the registrant requests additional uses, then additional studies might be required. In this example, assume the second use is for application to turf. Then, impacts on birds and aquatic organisms would be of concern, and therefore studies on those organisms would be necessary.

A similar situation exists for commercial chemicals under TSCA. A manufacturer may supply information to EPA that a substance is a site-limited intermediate with no significant releases to the environment. In this case, the exposure of concern would be exposure to workers and the appropriate test would be the mammalian reproduction test. If new uses were developed for the chemical which could result in environmental exposure, the ecological tests could be

required. The rationale for testing should not be limited solely to use considerations but should also consider potential releases from manufacturing, storage, transportation, and disposal.

e) Finally, as noted in Chapter Three, if a chemical is placed in the "hold box" based on the results of Tier 2 tests that are conducted on less than the full set of the five recommended taxa/tests, and the use and exposure patterns for the chemical substance subsequently change, EPA should determine whether the new use or uses warrant testing in one or more of the taxa that had previously not been included in Tier 2 testing of the chemical.

2. Determining When to Perform Alternative Tests

In Chapter Three the EDSTAC sets forth a definition of "function equivalency." In presenting the guidelines for determining when it may be appropriate to perform alternative Tier 2 tests, the EDSTAC wishes to emphasize the difference between alternative tests that meet the definition of functional equivalency from those that do not. As stated in Chapter Three, an assay, test, or endpoint can be considered to be "functionally equivalent" to a T1S or T2T assay, test, or endpoint when it provides equivalent information for each endpoint being studied. Furthermore, the EDSTAC-recommended assays, tests, and endpoints must be validated and standardized prior to EPA's use of functionally equivalent information.

The EDSTAC recommends that, as new tests are developed and proposed for inclusion in T2T, EPA should determine whether such tests are functionally equivalent to the recommended Tier 2 tests. If they are determined not to be functionally equivalent, the decision about whether and, if so, under what condition, any newly developed alternative tests should be used, should be subjected to the criteria listed below.

This section sets forth guidelines for determining when to conduct alternative tests, which are less comprehensive than the recommended tests, and therefore do not meet the definition of "functional equivalency."

The EDSTAC believes there are instances when a less comprehensive test for any of the five recommended taxa, when considered along with existing information, would provide sufficient information upon which to make a decision that meets the overall purpose of T2T (i.e., whether to go to the hold box or hazard assessment). The EDSTAC identified two alternative study designs for mammalian species. The EDSTAC believes such alternative designs could be developed for all species currently included in T2T. When and if such alternative tests are developed for the other taxa, in making the decision to perform a less comprehensive test EPA should consider the same issues, and apply the same principles that are described below for the alternative mammalian tests. The EDSTAC recommends alternative study designs be developed for other species/taxa only after their respective multi-generation study protocols are validated and standardized.

The primary considerations for determining whether an alternative, less comprehensive test should be conducted include an understanding of the toxicity profile of the chemical substance or mixture

under study; its mechanisms of action; exposure scenarios; use patterns; populations at risk; and other prior information. The criteria for determining when an alternative test might be run include: (1) a full two-generation reproductive toxicity study has been run in the past, but it either was conducted in accordance with the "old" guideline and/or the results in the previous study require additional follow-up that may be accomplished using one of the alternative protocols; (2) production volume and potential for exposure is low; or (3) there is low probability of at-risk populations being exposed. In applying criteria 2 and 3, the alternative test may be used more as a preliminary evaluation and, therefore, not necessarily as the last evaluation of the potential for EAT effects in a reproductive toxicity study. Furthermore, the EDSTAC wishes to note that chemical substances determined to be a high priority for the early phases of T1S based primarily on exposure considerations are not likely to meet these criteria; whereas the chemicals permitted to bypass T1S due to existing statutory requirements (i.e., food-use pesticides) will more often than not meet the first of these three criteria.

Below are potential scenarios for a specific chemical substance or mixture that requires a decision on whether to perform the two-generation mammalian test or one of the alternative tests. These examples pertain only to the mammalian reproduction tests as they are the only ones currently discussed in the report.

a) A chemical substance or mixture has not been through T1S, and thus, a weight-of-evidence evaluation cannot be performed. One may or may not have prior information which offers an understanding of the chemical substance or mixture's potential to show endocrine activity for the EAT hormone systems. In this case, the EDSTAC-recommended two-generation reproduction study would be conducted for each taxa/species which has been identified as being exposed or likely to be exposed. Whether or not all four of the non-mammalian tests would be performed should be determined in accordance with the principles described in Section V, C, 1 of this chapter.

b) A chemical substance or mixture has been through T1S and the weight-of-evidence evaluation concludes it does not have endocrine activity for EAT. The chemical substance or mixture has use patterns suggesting no or low exposure potential to populations of concern. In this case, no further screening or testing for EAT would be needed. However, a decision might be made to ask for a reproductive toxicity study to satisfy other regulatory requirements. One might argue that an alternative test protocol would provide sufficient information at this time since it is important to evaluate the potential for inducing reproductive effects of concern that may have other underlying mode(s)/mechanism(s) of action. Obviously, depending upon the outcome of the study and later discovery of additional relevant information, additional, follow-up studies could be required.

3. Determining Selection of Endpoints

In addition to guidance regarding selection of tests, the EDSTAC believes it is necessary to offer some guidance related to selection of which endpoints to include in performance of the tests for each chemical substance or mixture. Below are some potential scenarios regarding selection of which particular endpoints should be evaluated.

a) A chemical substance or mixture has been through T1S and the "weight-of-evidence" evaluation concludes it may have endocrine activity in all three hormone systems (E, A, and T). The chemical substance or mixture has use patterns suggesting exposure could possibly occur to populations of concern. In this case, the two-generation test, including the thyroid related enhancements (recommended later in the report), should be conducted.

b) A chemical substance or mixture has been through T1S and the "weight-of-evidence" evaluation concludes it may have endocrine activity for E and A, but not for T. The chemical substance or mixture has use patterns suggesting exposure could possibly occur to populations of concern. In this case, it would be appropriate to require the EDSTAC-recommended mammalian two-generation test, without the recommended enhancements for evaluating T. In other words, the study design would be tailored to provide results only for the endpoints of concern identified in T1S.

4. Chemicals That Bypass T1S and go Directly to T2T

In Chapter Three, the EDSTAC makes a distinction between two different scenarios under which a chemical would be permitted to bypass T1S and go directly to T2T. The first includes chemicals that have previously been subjected to two-generation mammalian and wildlife developmental and reproductive toxicity testing. The second includes chemicals for which the owner of the chemical has decided to voluntarily go to T2T without having completed the full T1S battery or any prior two-generation reproductive toxicity testing. The EDSTAC recommends elsewhere that both categories of "bypass chemicals" should be required to complete the HTPS assays.

T1S assays, in the aggregate, provide preliminary information on the presence of endocrine activity for EAT, the mechanism of action, and the species and sex at risk. The EDSTAC believes the absence of such information in the case of chemicals that voluntarily bypass T1S without having completed the full T1S battery or any prior two-generation reproductive toxicity testing provides sufficient justification for requiring such chemicals to complete all five tests in the T2T battery (i.e., the two-generation mammalian and non-mammalian tests with all the recommended endpoints). However, the determination of whether all of the non-mammalian tests would need to be conducted should be made consistent with the principles governing their selection, which were set forth in Section V, C, 1 of this chapter.

For chemicals that bypass T1S because they have previously been subjected to two-generation mammalian and wildlife developmental and reproductive toxicity testing, as noted above, the

EDSTAC believes these chemicals are the primary candidates for meeting the criteria specified for possible use of alternative Tier 2 tests. Once again, the determination of whether all of the non-mammalian tests would need to be conducted should be made consistent with the principles governing their selection, which were set forth in Section V, C, 1 of this chapter.

The EDSTAC recognizes it may be necessary, after completion of T2T, to conduct a limited number of assays that are similar, if not identical, to those that would have been conducted during T1S for chemicals which are permitted to bypass the T1S battery. The purpose of conducting these assays as part of T2T is to gain knowledge about specific mechanisms of action that are necessary to complete the hazard assessment step and/or to determine whether any adverse effects observed in T2T are in fact endocrine mediated.

5. Potential Need for Supplemental Information to Complete T2T

The EDSTP focuses on identifying agents that act as reproductive or developmental toxicants through the EAT endocrine-mediated mechanisms. Properly conducted, Tier 2 tests are intended to be the final arbiter of whether or not a substance is an endocrine disruptor for EAT. In other words, when the results of T2T are unambiguous, they provide a conclusive answer. Given the current definition of testing, both the expression of "endocrine-mediated" and "adverse effects" are necessary conditions for designation as an endocrine disruptor for EAT. In some cases, particularly for those chemicals that have received a positive in T1S, it will be apparent that the type of effect seen in T2T is endocrine-mediated. In other cases, particularly for those chemicals that bypass T1S, the results of T2T may not allow a judgment to be reached that the adverse effect is endocrine-mediated. The EDSTAC recommends that such information, if determined to be necessary to achieve a "weight-of-evidence" judgment, should be generated through further study using more focused assays, some of which may be drawn from the T1S battery.

Furthermore, it is inevitable with toxicological testing that equivocal results will sometimes be obtained in T2T. When this occurs one must look at the possible reasons for the ambiguity and see if there are ways to resolve it. Conducting other tests, or running assays to investigate the mechanism of action (if T1S was bypassed), may resolve the ambiguity and allow a more informed "weight-of-evidence" determination of the hazard potential of the chemical substance or mixture. Alternatively, repeating the study, perhaps at different exposure levels, may resolve the ambiguity.

D. Low Dose Considerations for T2T

1. Introduction to the Issue

Issues have been raised regarding approaches to regulatory toxicity testing that employ dosing regimens up to maximally tolerated dosages in order to identify a hazard and extrapolate from these doses to estimate risk or safe levels in the range of environmentally relevant exposures. There are two principal issues: (1) whether or not a threshold dose exists for receptor mediated toxicity; and (2) whether the dose response curve is monotonic or non-monotonic in nature.

With respect to the first issue, some believe that there is no threshold for effects of exogenous endocrine disruptors since there exist background levels of endogenous hormones that are already biologically active. Thus, any additional exposure constitutes an exceedance of the threshold (e.g., McLachlan et al., 1982; Wibbels et al., 1991; Russo, 1996; Gray et al., 1997b).

The second issue relates to the shape of the dose-response curve. Although monotonic curves may vary in slope, the slope of such curves is always in the same direction (either positive or negative including zero) and therefore there are no local maximum or minimum points along the curve. With non-monotonic dose-response curves, local minima can exist such that an effect may be pronounced at low doses, then becomes statistically insignificant or disappears at intermediate doses, and finally reappears (or a different effect appears) at higher doses. The issue raised by non-monotonic curves is that "high dose" testing may fail to detect toxicity that occurs in the "low dose" region of the dose-response curve. This is because the classical approach to finding the NOAEL by progressive reduction of the dose beginning at high doses will locate the nadir (i.e., a local minimum at some intermediate dose) but will not locate the second region of increased effect at doses below the apparent NOAEL. This issue is further complicated by the possibility of different effects at low doses as compared to high doses. There are examples of this phenomenon for endogenous hormones. For example, it is well known that testosterone stimulates sperm production up to a point, but at excess levels inhibits it. There is evidence that the developing mouse prostate responds in a non-monotonic manner to estrogens, in that prostate weight is increased initially, then decreased by higher maternal dosages of potent estrogens like DES and ethinyl estradiol. There is one report in the literature indicating that the weak estrogen bisphenol A has effects similar to these potent estrogens at low doses (Nagel et al., 1997). There is intense scientific debate surrounding these issues that centers on two principal questions: first, are data implicating xenobiotics in such phenomena reproducible and broadly generalizable to endocrine endpoints and endocrine active chemicals? and second, is the low dose phenomenon indicative of adverse effects at the individual or population level? If low dose phenomena are reproducible, generalizable, and related to adverse effects, the implications for regulatory toxicity testing and risk assessment are profound. The EDSTAC recognizes there are divergent scientific opinions on the "low dose" issue at the present time and that more research is necessary to answer these questions.

The EDSTAC notes that, historically, testing has sometimes missed critical endpoints either by: (1) failing to dose during the most sensitive life stage (Morrissey et al., 1987); (2) failing to test in a susceptible organism (Chamberlin, 1979; Fraser, 1988); or (3) failing to examine subtle (yet biologically important) endpoints. For example, early studies on the developmental effects of PCBs in rodents identified fetotoxicity as a critical endpoint, yet these studies failed to test at low enough doses or to measure subtle enough endpoints. As a result these early studies missed the neurotoxic effects of PCBs which occur at much lower doses (Tilson et al., 1990).

These omissions in testing may lead either to missing a critical effect completely, or to identifying a NOAEL which is erroneously high. The EDSTAC has attempted to minimize the likelihood of these types of errors by requiring testing in a variety of organisms during sensitive life stages. A variety of endpoints which appear to be low-dose sensitive have also been added to the EDSTP

testing protocols. In addition, a number of endocrine disruptor-sensitive endpoints have already been added to EPA's reproductive toxicity testing guidelines (OPPTS Public Draft, 1996; U.S. EPA TSCA finalized testing guidelines, 1997; U.S. EPA FIFRA finalized testing guidelines, 1998).

The Committee agrees, however, that dose selection in T2T must include special attention to setting the low dose. In particular, the low dose should not be selected by identifying the high dose and then dropping the dose by a fixed formula of a couple of orders of magnitude. Instead, a number of considerations need to go into selecting the low dose, including the results of prior information (e.g., HTPS, toxicity testing, pharmacokinetic, and epidemiology data). Information about environmental exposure levels might also be used where appropriate. Finally, range-finding studies are currently constructed so as to inform both the high dose and the low dose for the definitive testing, with inclusion of low-dose sensitive endpoints in the range-finding study. Range-finding studies, with fewer animals per group and a more limited set of endpoints than in definitive tests, will not necessarily identify the appropriate low dose or a NOAEL but they will indicate whether effects are observed for low dose sensitive endpoints and therefore whether low doses should be included in the definitive tests. These precautions will minimize the likelihood that critical effects will be missed or that erroneously high NOAELs will be identified in T2T.

2. Recommended Project to Address Low Dose Issues

As previously discussed, questions have been raised as to the adequacy of conventional toxicology study designs for assessment of endocrine active substances, particularly with regard to low dose selection and the identification of no-observed-adverse-effect-levels (NOAELs). The EDSTAC believes a project is required to resolve the underlying uncertainties and controversy about these issues. The purpose of the project is to address the nature of the dose-response curves for exogenous estrogenic substances in order to allow more informed judgments about appropriate toxicology study designs for substances that have hormonal activity. The recommended project focuses only on substances with estrogenic activity, since it has been reported that some estrogens can produce nonmonotonic dose response curves at environmentally relevant dosage levels. However, the results of the study could be more broadly applicable to substances with other types of hormonal activity (e.g., those that have androgenic or anti-androgenic activity).

Several very important studies related to low dose issues are currently in progress. The exact design of the project (e.g., chemicals to be tested, strains of animals, endpoints to be evaluated) will be based on a thorough evaluation of all relevant data that can be made available. The EDSTAC believes this evaluation should be completed within a reasonable timeframe, such as four to six months. Preliminary discussions to date have focused on a two-phased project. Phase I (Replication Studies) is intended to replicate previously reported low dose results in male and female mice with one test substance and a positive control substance (e.g., DES). The replication studies would be done in at least 3 different laboratories using an identical, mutually agreed upon protocol. The Phase I studies will allow a systematic evaluation of inter-laboratory variability in the assessment of various estrogen-responsive endpoints in male mice. Depending on the timing for completing studies currently in progress, or results of studies, if they have been completed,

Phase I may not be required. If Phase I is required, it should be conducted simultaneously with Phase II (described below).

Phase II (Exploratory Studies) is intended to evaluate potential new endpoints and compare the effects in male and female rats and mice with approximately four to seven different substances, with varying degrees of estrogenic potency (including estradiol and at least one natural estrogenic substance) using an identical protocol for each substance. The number of substances, as well as the number of laboratories, will depend on available resources. The Phase II studies will allow a systematic determination of any species or sex differences in sensitivity, as well as elucidation of any qualitative differences in the responses to the various test substances. Phase II will allow assessment of the complete time course for any observed effects, and a determination of whether or not observations early in life are predictive of any specific adverse outcomes during later life stages.

If the results of the project confirm there is a need to include additional low doses for Tier 2 testing of substances with estrogen, androgen, or thyroid activity, there are implications for additional testing of any substances that may already have been subjected to dosing in the EDSTAC-recommended Tier 2 tests. For any substance that has previously been tested and found to be positive in T2T (i.e., produces adverse effects on EAT sensitive endpoints), it will be necessary to retest that substance with additional low doses. This additional low dose testing would be done for the purpose of conclusively identifying the NOAEL.

If the results of the project demonstrate the need to include any additional EAT endpoints, there would be a need for retesting to assess those new endpoints for any substances previously tested using the EDSTAC-recommended T2T protocols. Additional endpoints would be those that are plausibly linked to adverse effects and not covered in current tests, or those that would be more sensitive than those in current tests.

Given the current state of the science, the EDSTAC believes its recommended T2T protocols (once standardized and validated) will be adequate to determine whether chemicals exhibit EAT-mediated adverse effects. If the results of the project confirm there is a need to include additional low doses, but there is not a need to include additional EAT endpoints, then it will be necessary to retest chemicals that have been found positive in the EDSTAC recommended T2T protocols to conclusively identify the NOAEL, if one exists. However, it will not be necessary to retest chemicals that have already been found negative, and placed in the "hold box," using the EDSTAC recommended T2T protocols. As indicated in Chapter Three, all chemicals placed in the "hold box," can be reconsidered for additional testing if "new information on the endocrine disrupting potential of the chemical substance or mixture becomes available and it is determined that this new information warrants additional testing."

If retesting is required, a test concentration that produced adverse effects in the EDSTAC-recommended T2T protocol (if any) would be required as the high dose of the repeat study to demonstrate reproducibility of the initial finding.

If the project demonstrates that conventional protocols are adequate for identifying the NOAEL of endocrine active substances, then the NOAEL derived from conventional T2T should not be contested.

The EDSTAC recommends that a collaborative group, involving government, industry, and appropriate individuals in academia, design the study protocols, be kept abreast of the conduct of the studies, evaluate results, and develop overall conclusions and recommendations.

E. Methods to Select the Target Doses for T2T

For T2T in mammals, other vertebrates, and invertebrates, the EDSTAC recommends that the information to select doses used in the performance of these tests include:

1. existing information such as that available during the priority setting phase including results from the HTPS (or its equivalent by bench-level assays);
2. results from T1S (including the range-finding study results);
3. results from other assays or tests for pesticide registration, etc.; and
4. results from range-finding studies.

Range-finding studies specifically for T2T should be performed at multiple doses (at least five) with a limited number of animals per dose, an abbreviated duration (which must include exposures during gestation or egg development and lactation), and a limited number of relevant endpoints including low dose sensitive ones. If further research validates the low dose concern, the EDSTAC would recommend using the low dose sensitive endpoints in the range-finding study to determine the need for inclusion of low doses in the definitive T2T. Range-finding studies are already conducted for toxicity studies, and have the principal purpose of defining both the maximal level that can be dosed and, at the other end of the dose range, informing a dose that will have no observable effect on the most sensitive endpoint in the full-scale study. Endpoints, identified in recent publications, which, at present, appear to be low dose sensitive include: prostate weight (for mammals), epididymal sperm concentration (for mammals), other accessory sex organ weights (all vertebrates), thyroid weight (all vertebrates), reproductive capability (all T2T), and vaginal threads (for mammals). All of these, except for vaginal threads, are included in the 1996 guidelines; however, vaginal threads would be identified during the examination of offspring females for vaginal patency. New and/or different low dose sensitive endpoints may be identified as new data are generated.

Current toxicological test guidelines generally require testing at a minimum of three dose levels plus a control. These guidelines specify that the top dose level should be a maximally tolerated dose (MTD), that is, a dose which by definition is toxic but which does not result in excessive mortality (not to exceed 10%). In reproductive and developmental toxicity studies, the MTD is usually based on parental or maternal toxicity, which is expressed as depressed body weight gain, actual weight loss, reduced feed and/or water consumption, treatment-related clinical signs of toxicity, etc. The MTD is based on available toxicity information such as data from a range finding study. The next lower dose is ideally set at an intermediate toxic dose and the lowest dose

at a level at which no toxic effects are observed. If additional lower doses are included in the study because of identified concerns at low doses, the additional doses should be widely spaced (perhaps in orders of magnitude) to identify the nature of the dose-response curve in the low dose region. The toxicity upon which the MTD is based may or may not be related to the endpoints that are the object of the investigation (e.g., cancer, neurotoxicity, reproductive, or developmental effects).

The rationale and results of range-finding studies should be included in the submission of the T2T results for evaluation by the Agency.

F. Testing Antithyroid Activities in T2T

Thyroid hormones are well known to play essential roles in vertebrate development (Dussault and Ruel, 1987; Myant, 1971; Porterfield and Hendrich, 1993; Porterfield and Stein, 1994; Timiras and Nzekwe, 1989). Experimental work focused on the effects of thyroid hormone on brain development in the neonatal rat supports the concept of a "critical period," during which thyroid hormone must be present to avoid irreversible damage (Timiras and Nzekwe, 1989). Though the duration of this critical period may be different for different thyroid hormone effects, the general view has developed that this is the period of maximal developmental sensitivity to thyroid hormone, and it occurs during the lactational period in the rat (Oppenheimer et al., 1994; Timiras and Nzekwe, 1989). Although thyroid hormone receptors are expressed in fetal rat brains (Bradley et al., 1989; Strait et al., 1990) and thyroid hormone can exert effects on the fetal brain (Escobar et al., 1990; Escobar et al., 1987; Escobar et al., 1988; Porterfield, 1994; Porterfield and Hendrich, 1992; Porterfield and Hendrich, 1993; Porterfield and Stein, 1994), the lactational period represents a stage of rapid expansion of the central nervous system that coincides with a large increase in the expression of thyroid hormone receptors (Perez-Castillo et al., 1985) and an increase in the number of demonstrated effects of thyroid hormone on brain development.

In conducting thyroid-related tests in Tier 2, the EDSTAC recommends using an approach in which dosing occurs during the fetal and lactational period. In addition, there are a variety of endpoints that would provide reliable markers of thyroid disruption. Brain weight offers a simple measure, though it is not thyroid specific. Characteristics of myelination, or of myelin basic protein expression (either mRNA or protein), would provide a more selective measure (Bhat et al., 1981; Bhat et al., 1979; Farsetti et al., 1991; Figueiredo et al., 1993; Rodriguez-Pena et al., 1993; Shanker et al., 1987). In this regard, the expression of myelin basic protein and/or neurogranin/RC3 may offer the simplest and most specific endpoints of thyroid disruption during the perinatal period (Farsetti et al., 1991; Iniguez et al., 1993). These mRNAs are both enormously abundant and robustly affected by thyroid hormone. However, their sensitivity to xenobiotics has not been studied. A list of existing endpoints for thyroid hormone function, and additional ones recommended by the EDSTAC for validation and inclusion, are found in Table 5.3.

VI. Recommended Tier 2 Testing Battery

A. Outline of Recommended T2T Battery

The EDSTAC recommends that the T2T battery include a mammalian two-generation reproductive toxicity study, or a less comprehensive test in accordance with the guidelines outlined above, and tests addressing four additional taxonomic groups, including birds, amphibians, fish, and invertebrates as follows:

Mammalian Tests

1. Two-Generation Mammalian Reproductive Toxicity Study; or
2. A Less Comprehensive Test:
    a) Alternative Mammalian Reproductive Test; or
    b) One-Generation Test.

Multigeneration Tests in Other Taxa

1. Avian Reproduction (with bobwhite quail and mallard)
2. Fish Life Cycle (fathead minnow)
3. Mysid Life Cycle (*Americamysis*)
4. Amphibian Development and Reproduction (*Xenopus*)

B. Two-Generation Mammalian Reproductive Toxicity Study

The two-generation reproductive toxicity study in rats (TSCA 799.9380, August 15, 1997; OPPTS 870.3800, Public Draft, February, 1996; OECD no. 416, 1983; FIFRA, Subdivision F, Guidelines 83-4) is designed to comprehensively evaluate the effects of a chemical on gonadal function, estrous cycles, mating behavior, fertilization, implantation, pregnancy, parturition, lactation, weaning, and the offspring' ability to achieve adulthood and successfully reproduce, through two generations, one litter per generation. Administration is usually oral (dosed feed, dosed water, or gavage) but other routes are acceptable with justification (e.g., inhalation). In addition, the study also provides information about neonatal survival, growth, development, and preliminary data on possible teratogenesis. The experimental design for a two-generation reproductive toxicity study is presented in Figure Q.1, which is found in Appendix Q, Tier 2 Test Study Designs.

In the existing two-generation reproductive toxicity test, a minimum of three treatment levels and a concurrent control group are required. At least 20 males and sufficient females to produce 20 pregnant females must be used in each group as prescribed in this current guideline. The highest dose must induce toxicity but not exceed 10% mortality. In this study, potential hormonal effects can be detected through behavioral changes, ability to become pregnant, duration of gestation, signs of difficult or prolonged parturition, apparent sex ratio (as ascertained by anogenital distances) of the offspring, feminization or masculinization of offspring, number of pups, stillbirths, gross pathology and histopathology of the vagina, uterus, ovaries, testis, epididymis,

seminal vesicles, prostate, and any other identified target organs, and gynecological assesments (evaluation of parental estrous cyclicity, onset of puberty, acquisition of VO).

Table 5.3 provides a summary of the endpoints evaluated within the framework of the experimental design of the updated two-generation reproductive toxicity test (and some recommended additional endpoints for validation and inclusion to cover EAT concerns). These endpoints are comprehensive and cover every phase of reproduction and development. Tests that measure only a single dimension or component of hormonal activity, (e.g., *in vitro* or short-term assays) provide supplementary and/or mechanistic information, but cannot provide the breadth of information listed in Table 5.3, which is critical for risk assessment.

Additionally, in this study type, hormonally induced effects such as abortion, resorption, or premature delivery as well as abnormalities and anomalies such as masculinization of the female offspring or feminization of male offspring, can be detected. Substances such as the phytoestrogen, coumesterol, and the antiandrogen, cyproterone acetate, which possess the potential to alter normal sexual differentiation, were similarly detected in this study test system (i.e., 1982 Guideline). The initial prebreed exposure period (10 weeks) of the two-generation reproductive toxicity test also provides information on subchronic exposures which can be used for other regulatory purposes.

C. Alternative Approaches to Mammalian T2T

The EDSTAC acknowledges that the developing organism may be uniquely sensitive to the effects of endocrine-active agents. Therefore, any mammalian Tier 2 test should include a careful assessment of the consequences of *in utero* and lactational exposure on subsequent growth and development. Below are two alternative mammalian tests the EDSTAC recommends be included, once validated and standardized, as part of T2T. The determination on whether to perform the two-generation test or one of the alternative tests should be made consistent with the criteria described in Chapter Five, Section V, C, 2.

Although EAT-relevant endpoints are not optimally detected by developmental toxicity study designs, both of these alternative tests can be modified to allow assessment of term fetuses by adding additional F0 females to each dose group (i.e., create satellite groups) and terminating these satellite females just prior to anticipated parturition (i.e., on gd 20-21) and performing gestational and fetal structural evaluations (i.e., ovarian corpora lutea, uterine implantation sites,

Table 5.3

Mammalian Tier 2 Test Endpoints

Below are two types of lists. First, those endpoints required in current EPA test guidelines 1996. Second, additional endpoints recommended by EDSTAC for validation and inclusion in both the recommended two-generation test, as well as the alternative mammalian tests discussed below. These additional endpoints will detect estrogen, androgen, and thyroid hormone perturbations.

As discussed above, in Chapter Five, Section V, C, the default assumption is that all of these endpoints would be evaluated unless the conditions which are set forth in the guidelines for determining the selection of endpoints are met.

Current Guideline Endpoints Sensitive to Estrogens/Antiestrogens

- sexual differentiation
- gonad development (size, morphology, weight) > accessory sex organ (ASO) development
- ASO weight ± fluid; histology
- sexual development and maturation: acquisition of vaginal patency (VP), preputial separation (PPS)
- fertility
- fecundity
- time to mating
- mating and sexual behavior
- ovulation
- estrous cyclicity
- gestation length
- abortion
- premature delivery
- dystocia
- spermatogenesis
- epididymal sperm numbers and morphology; testicular spermatid head counts; daily sperm production (DSP); efficiency of DSP
- gross and histopathology of reproductive tissues
- anomalies of the genital tract
- viability of the conceptus *in utero* (prenatal demise)
- survival and growth of offspring
- maternal lactational behaviors (e.g., nursing, pup retrieval, etc.)

Recommended Additional Endpoints for Validation and Inclusion

- accessory sex organ function (secretory products)
- sexual development and maturation (nipple development and retention)
- androgen and estrogen levels
- LH and FSH levels
- testis descent

Current Guideline Endpoints Sensitive to Androgens/Antiandrogens

- altered apparent sex ratio (based on AGD)
- malformations of the urogenital system
- altered sexual behavior
- changes in testis and accessory sex organ weights
- effects on sperm numbers, morphology, etc.
- retained nipples in male offspring
- altered AGD (now triggered from PPS/VP)
- reproductive development; PPS/VP (puberty)
- male fertility
- agenesis of prostate
- changes in androgen-dependent tissues in pups and adults (not limited to sex accessory glands)

Current Guideline Endpoints Sensitive to Thyroid Hormone Agonists/Antagonists (general)

- growth, body weight
- food consumption, food efficiency
- developmental abnormalities
- perinatal mortality
- testis size and DSP
- VP; PPS

Recommended Additional Endpoints for Validation and Inclusion

- neurobehavioral deficits (see developmental landmarks below)
- TSH, T4, thyroid weight and histology (e.g., goiter)
- developmental landmarks:
    - prewean includes pinna detachment, surface righting reflex, eye opening, acquisition of auditory startle, negative geotaxis, mid-air righting reflex, motor activity on PND 13, 21, etc.
    - postwean includes motor activity PND 21 and postpuberty ages (sex difference); learning and memory PND 60 - active avoidance/water maze
    - brain weight (absolute), whole and cerebellum
    - brain histology

total, resorbed, dead and live fetuses, live fetal number, sex, weight, external, visceral, and skeletal alterations), i.e., a "standard" developmental toxicity evaluation by OPPTS 1996 draft guidelines; USEPA TSCA guidelines 870.3700, 1997; FDA guidelines, 1993. The remaining females in each group would continue on study as described below.

1. Alternative Mammalian Reproduction Test

A graphical representation of the study design (Figure Q.2) and additional descriptive text for the Alternative Mammalian Reproduction Test (AMRT) are provided in Appendix Q, Tier 2 Test Study Designs. The objectives of this test are to describe the consequences of *in utero* and/or lactational exposure on reproduction and development from compounds that displayed EAT activity in the T1S. If validated, this test may be used, under certain defined circumstances, instead of the recommended two-generation reproductive toxicity test (TSCA guidelines, 1997) in T2T. In this regard it will be conducted with at least three treatment groups plus a control and include endpoints sensitive to chemicals that alter development via EAT activities. As with the two-generation mammalian reproductive toxicity study, the default assumption is that all of the endpoints in Table 5.3 would be evaluated in the AMRT, unless the conditions set forth in the guidelines for determining the selection of endpoints are met.

The AMRT involves exposure of maternal rats (designated F0 generation) from gestational day 6 (time of implantation), through parturition (birth), and through the lactation period until weaning of offspring (designated F1 generation) on postnatal day 21. F1 offspring (both sexes) are retained after weaning with no exposures for 10 weeks and then mated within groups. F1 males are necropsied after the mating. F1 females and their litters (designated the F2 generation) are retained until the F2 generation is weaned. F0 females (and a subset of F1 weanlings) are necropsied with organ weights and possible histopathology. F1 animals are evaluated for reproductive development (VP, PPS), estrous cyclicity, and, at necropsy, for organ weights, possible histopathology, andrological assessments, and T3/T4 (with TSH triggered). F2 weanlings are counted, sexed, weighed, examined externally, and discarded.

The AMRT differs from the "standard" two-generation study design in that it:

a) does not include exposures prior to mating, during mating, or during the early preimplantation stage of pregnancy in the dams;
b) does not include exposures to parental males; and
c) does not include direct exposure to the postweanling offspring; potential exposure is limited to *in utero* transplacental and/or lactational routes.

The AMRT differs from the one-generation test (see below) in that its study design provides for:

a) exposure to the F0 dam only from gd6, through weaning of the F1 offspring on pnd 21;
b) no exposure to parental males;
c) mating of the F1 animals (who have not been directly exposed) to produce F2 offspring; and
d) following the F2 offspring to weaning on pnd 21.

2. One-Generation Test

A second alternative to the standard two-generation reproductive toxicity test is a one-generation reproductive toxicity test. A graphical representation of the one-generation test (Figure Q.3), and additional text, is provided in Appendix Q, Tier 2 Test Study Designs. It has been used in rats and mice. It has been used as a range-finding study prior to performance of a guideline two-generation (or more) study for the last 10 years under EPA (TSCA/FIFRA) GLPs; the design is similar to that used by Sharpe et al., 1996. This is a shortened, scaled-down version of the new draft OPPTS and Final TSCA guidelines for reproductive toxicity testing. As with the two-generation mammalian reproductive toxicity study, the default assumption is that all of the endpoints in Table 5.3 would be evaluated in the one-generation test, unless the conditions set forth in the guidelines for determining the selection of endpoints are met.

The one-generation test is a less comprehensive evaluation of functional reproductive development than the AMRT (since it does not follow F1 animals through production of F2 offspring), but it has the advantage of assessing postnatal development and adult reproductive capacity after *in utero* lactational and post-lactational exposure. In the presence of continued exposure, the postnatal component of the test is extended to evaluate acquisition of vaginal patency, preputial separation, estrous cyclicity, and andrological assessments in the F1 offspring. Inappropriate retention of Mullerian duct derivations (e.g., oviducts) in males and of Wolffian duct derivatives (e.g., seminal vesicles, epididymides) in females can be identified in all three proposed tests (with or without satellite F0 females and examination of term fetuses).

The one-generation test involves a short prebreed exposure period for male and female rats of the initial parental generation (designated F0), and exposure continues through mating, gestation, and lactation of F1 litters. F0 males are necropsied after F1 deliveries; F0 females are necropsied after F1 weaning. Postweanling F1 animals are directly exposed for a 10-week postwean period and are then necropsied. F1 animals are evaluated for reproductive development (VP, PPS), estrous cyclicity and at necropsy for organ weights, possible histopathology, andrological assessments, and T3/T4 (TSH triggered). F0 animals will undergo the same necropsy assessments.

The one-generation test differs from the “standard” two-generation study design in that it:

a) is shorter (basic design calls for two weeks, can be prolonged) than in the standard two-generation study (10 weeks to encompass one full spermatogenic cycle in rats), though it does include a prebreed exposure period; and
b) does not evaluate effects of *in utero* and/or lactational exposure (and beyond) on generation of F2 offspring though it does include direct exposure of F1 offspring after weaning, including exposure through puberty and sexual maturation. F1 male and female reproductive organs (weight/histology), estrous cyclicity, and andrological endpoints are assessed at scheduled necropsy on PND 90 ± 2.

The one-generation test differs from the AMRT in that its study design provides for:

a) exposure to both male and female F0 parental animals prior to mating, during mating, and during gestation and lactation of F1 offspring (F0 males are necropsied after F1 deliveries, F0 females are necropsied after F1 weaning);
b) direct exposure of postweanling F1 offspring after lactation until termination; and
c) no mating of F1 animals to produce F2 offspring.

D. Description of the Tests for Other Animal Taxa

The EDSTAC agrees T2T should address at least four other animal taxonomic groups, including birds, amphibians, fish, and invertebrates. Each of the four basic non-mammalian tests should be multigenerational, as is the basic mammalian reproduction study. It is recommended that the following standardized tests be used as a basis for a non-mammalian battery:

1. Avian Reproduction (with bobwhite quail and mallard)
2. Fish Life Cycle (fathead minnow)
3. Mysid Life Cycle (Americamysis)
4. Amphibian Development and Reproduction (Xenopus)

Except for the amphibian study, these tests are routinely performed for chemicals with widespread outdoor exposures that are expected to affect reproduction. Modifications to each may be warranted to enhance the ability to detect endocrine-related effects. The amphibian test, though not standardized, is considered warranted because of the extensive fundamental knowledge base on amphibian development and reproduction.

Just as for mammalian testing, there may be instances when less comprehensive study designs would be adequate. Considerations for determining whether the full battery of comprehensive non-mammalian tests should be implemented include an understanding of mechanisms of action, environmental fate and transport, persistence, potential for bioaccumulation, and potential ecosystems exposed.

Production volume is also a consideration for less comprehensive approaches. Comprehensive assessments of environmental toxicity, including chronic toxicity assays in a variety of species, are already generated for pesticides and very high production volume chemicals (>1,000 tons per annum) in Europe. The European Union explicitly requires less comprehensive assessments for lower production volume chemicals, with additional testing required as production increases. As with mammalian assessments, these moving triggers recognize that the potential for exposure is correlated with production volume. While there are no explicitly required data sets in the U.S. under TSCA, similar decisions are made on data adequacy based on other information, including production volume.

There are a number of alternative, less comprehensive assays that may be appropriate to consider for environmental toxicity assessments. These might include shorter-term avian development tests, and the fish early life cycle test and Daphnia reproduction test, both of which are already established protocols.

1. Avian Reproduction Test

While birds are not included as subjects in the T1S battery, it is important to evaluate the effects of exposure of birds to chemical substances or mixtures with endocrine activity. Furthermore, birds are fundamentally different from mammals in the control of sexual differentiation (males are the homogametic sex) so results using mammalian subjects will not provide complete information relevant to birds.

Use of the EPA's Avian Reproduction Test guidelines (OPPTS 850.2300) is recommended, modified to include the additional endpoints presented below to make the test more sensitive to chemical substances or mixtures with endocrine activity. Table 5.4 provides a summary of the endpoints evaluated within the framework of the Avian Reproduction Test (and recommended additional endpoints for validation and inclusion to cover EAT concerns). Two important extensions of this guideline are recommended: (1) modification and standardization of the husbandry and dosing of the offspring from EPA's Avian Reproduction Test guidelines (OPPTS 850.2300) to create a two-generation avian reproduction test; and (2) using the procedures of the modified Avian Reproduction Test protocol, evaluate an additional exposure pathway (i.e., direct topical exposure, which is common in the wild, by dipping eggs). The recommended extensions to the guideline are outlined in Appendix Q.

In the current Avian Reproduction Test guidelines, two species are commonly used, mallards and northern bobwhite. Exposure of adults begins prior to the onset of maturation and egg laying and continues through the egg-laying period; their offspring are exposed, in early development, by material deposited into the egg yolk by the females. These offspring can be used efficiently to test for the effects of chemical substances or mixtures on avian development. There are several endpoints currently required (see OPPTS 850.2300, c, 2) that are particularly relevant to disruption of endocrine activity, including: eggs laid, cracked eggs, eggshell thickness, viable embryos, and chicks surviving to 14 days. The guidelines should be extended with additional observations made for circulating steroid titers, thyroid hormones, major organ (including brain) weights, gland weights, bone development, leg and wing bone lengths, ratios of organ weights to bone measurements, skeletal x-ray, histopathology, functional tests, and reproductive capability of offspring (Baxter et al., 1969; Bellabarba et al., 1988; Dahlgren and Linder, 1971; Emlen, 1963; Cruickhank and Sim, 1986; Fleming et al., 1985a; Fleming et al., 1985b; Fox, 1976; Fox et al., 1978; Freeman and Vince, 1974; Hoffman and Eastin, 1981; Hoffman and Albers, 1984; Hoffman, 1990; Hoffman et al., 1993; Hoffman et al., 1996; Jefferies and Parslow, 1976; Kubiak et al., 1989; Maguire and Williams, 1987; Martin, 1990; Martin and Solomon, 1991; McArthur et al., 1983; McNabb, 1988; Moccia et al., 1986; Rattner et al., 1982; Rattner et al., 1987; Summer et al., 1996; Tori and Mayer, 1981). Other avian assays were considered including the Japanese quail androgenic assay (proctodeal gland), egg injection, draft OECD Japanese quail

reproduction, and two generation avian reproduction tests, but were not selected because the endpoints addressed were limited or there was a lack of accepted and standardized methods.

2. Fish Life Cycle Test

The freshwater fathead minnow *Pimephales promelas* is the recommended species to be used and is continuously exposed from fertilization through development, maturation, reproduction, and early development of offspring with a test duration of up to 300 days. The fathead minnow is also the recommended species for use in the screening battery for the fish gonadal recrudescence assay, and as such, the relevance of any activity detected in the screening assay would be evaluated. However, EDSTAC recommends a performance-based approach to species selection and, as more appropriate species are developed and validated, EDSTAC strongly encourages their use. For example, if exposure to a particular chemical substance or mixture is predominantly estuarine or marine, the estuarine sheepshead minnow *Cyprinodon variegatus* may be substituted since experience and an established method exist for this species.

Likewise, other species may offer more specificity for certain endpoints or geographic relevance and should be considered in a performance-based approach. For example, the sex-linked color gene in a medaka strain may afford early, non-intrusive specificity for determining the genetic sex of test animals following estrogenic and androgenic screens or tests and anadromous salmon may have better value to examine thyroid function in a sea-water challenge (smoltification) assay.

Fish are the most diverse and least homologous to mammals of all vertebrates. Reproductive strategies extend from oviparity, to ovoviviparity, to true viviparity. The consequences of an endocrine disruptor may be quite different across the many families of fishes. As a first step though, only a fathead minnow, or in special cases the sheepshead minnow, life cycle test is suggested to confirm and quantify any effects detected by the Tier 1 battery. Subsequent tests with other species will then be a function of the risk assessment and nature of the hormones involved and effects expected/obtained.

The fish life cycle test (OPPTS 850.1500) follows procedures outlined in Benoit, 1981, for the fathead minnow, and Hansen et al., 1978, for the sheepshead minnow. In general, the test begins with 200 embryos distributed among eight incubation cups in each treatment group. When hatching is completed, the number of larvae are reduced to 25 individuals, if available, which are released to each of four replicate larval growth chambers. Four weeks following their release into the larval growth chambers, the number of juvenile fish are reduced again and 25 individuals, if available, are distributed to each of two replicate adult test chambers. When fish reach sexual maturity, fish are separated into spawning groups (pairs or one male/two females) with a minimum of eight breeding females. Remaining adults will be maintained in the tank but will be segregated from the spawning groups. Adults will be allowed to reproduce, at will, until the 300th day of exposure. Alternatively, the test may be continued past 300 days until one week passes in which no eggs from any group have been laid. The embryos and fish are exposed to a geometric series of at least five test concentrations, a negative (dilution water) control, and, if necessary, a solvent control.

Table 5.4

Avian Reproduction Test Endpoints

Current Guideline Endpoints Sensitive to Estrogens/Antiestrogens, Androgens/ Antiandrogens, and/or Hypothalamic-Pituitary-Gonadal Axis

- egg production
- eggs cracked
- viable embryos (fertility)
- eggshell thickness
- fertilization success
- live 18-day embryos
- hatchability
- 14-day-old survivors

Recommended Additional Endpoints for Validation and Inclusion

- sex ratio
- major organ (including brain) weights
- gland weights
- histopathology
- plasma steroid concentrations
- neurobehavioral test (e.g., nest attentiveness)

Current Guideline Endpoints Sensitive to Thyroid Hormone Agonists/Antagonists

- body weight of adults
- food consumption of adults
- body weight of 14-day-old survivors
- developmental abnormalities

Recommended Additional Endpoints for Validation and Inclusion

- plasma T3/T4
- thyroid histology
- bone development (skeletal x-ray)
- ratio of organ weights to bone measurements
- neurobehavioral test (cliff test)
- cold stress test

Assessment of effects on offspring of the parental group (first filial or F1 generation) will be made by collecting two groups of 50 embryos from each experimental group and incubating those embryos. When embryos hatch, the number of larvae hatched from each group will be impartially reduced to 25, if available, and released into the larval growth chambers. After four weeks of exposure, lengths and weights of surviving individuals will be made.

Observations are made of the effects of the test substance on embryo hatching success, larvae-juvenile-adult survival, growth of parental and F1 generation, and reproduction of the adults. Table 5.5 provides a summary of the endpoints evaluated within the framework of the Fish Life Cycle Test (and recommended additional endpoints for validation and inclusion to cover EAT concerns).

3. Mysid Life Cycle Test

Invertebrates (especially arthropods such as insects and crustaceans) constitute the vast majority of animal species on earth. Yet, relatively few invertebrate toxicity test protocols are routinely used in regulatory toxicity testing, and none have been designed with endocrine endpoints in mind. Nevertheless, invertebrate growth, reproduction, and development are under endocrine control. However, invertebrate endocrine systems and hormones are not directly analogous to those of vertebrates.

Two invertebrate life cycle toxicity tests are commonly used in chemical and pesticide testing, both using crustaceans. The opossum or mysid (Order Mysidacea) shrimp is an estuarine species, whereas water fleas or daphnids (Daphnia magna or Daphnia pulex) are freshwater species. The former is sexually dimorphic with males and females, whereas the latter undergoes parthenogenetic reproduction for the majority of its life cycle. Although sexual reproduction occurs in Daphnia, the standard test protocol (U.S. EPA Public Draft OPPTS 850.1300 and OECD 202) is designed solely for the parthenogenetic reproductive phase. The chronic tests for both species are designed to provide No Observed Effect Levels (NOEL) using apical effect endpoints for fecundity and growth.

Although neither chronic test was designed to examine endocrine specific endpoints, both species are crustaceans and therefore share common physiology. Ecdysone is a steroid hormone that regulates growth and molting in arthropods, and exhibits some functional and structural similarities to estrogen. The central role of ecdysone makes it an attractive candidate for examining endocrine effects in invertebrates; however, other possibilities also exist. Morphogenetic and reproductive development of arthropods is controlled in part by juvenile hormone (JH). Methyl farnesoate is a JH like compound that may play a role in reproduction and development (Borstet et al., 1987; Laufer et al., 1987a,b).

Invertebrate hormones are beyond the immediate scope of the EDSTAC which has focused on the vertebrate estrogen, androgen, and thyroid hormones. Nevertheless, invertebrate hormones that are functionally equivalent to estrogen, androgen, and thyroid need to be examined in more

Table 5.5

Fish Life Cycle Test Endpoints

Current Guideline Endpoints Sensitive to Estrogens/Antiestrogens, Androgens/Antiandrogens, and/or Hypothalamic-Pituitary-Gonadal Axis

- viability of embryos
- time to hatch
- spawning frequency
- egg production
- fertilization success

Recommended Additional Endpoints for Validation and Inclusion

- sexual differentiation (tubercle formation, gonadal histology)
- sex ratio
- gonadosomatic index
- gamete maturation (production, final oocyte maturation, sperm motility test, etc.)
- vitellogenin
- plasma steroid concentrations
- *in vitro* gonadal steroidogenesis

Current Guideline Endpoints Sensitive to Thyroid Hormone Agonists/Antagonists

- growth, length, and body weight
- developmental abnormalities

Recommended Additional Endpoints for Validation and Inclusion

- plasma T3/T4
- neurobehavioral tests (e.g., activity level and swimming performance, nesting (spawning) behavioral endpoints such as territory defense, courtship, and egg protection and care, or whether test fish retain the ability to avoid known deterrent chemicals after exposure to a test chemical)
- thyroid histology

depth. More importantly, chemicals that affect these vertebrate hormones may also affect invertebrate hormones resulting in altered reproduction, development, and growth.

Chemicals with estrogenic properties are reported to have altered normal function of ecdysone systems (Mortimer, 1993; 1994; 1995a; 1995b; Chu et al., 1997). (Satyanarayana et al., 1994) showed stimulation of vitellogenin in insect prepupae and pupae by methoprene, a JH mimic with retinoid properties. Whether vitellogenin production is controlled through an estrogen receptor or an alternative mechanism is not crucial for obtaining test results that show alteration occurs.

Therefore, the mysid shrimp chronic life cycle test (OPPTS 850.1350) may be adapted to determine whether chemicals that affect hormonal activity in vertebrates also affect arthropods. Once adapted to include reproductive and developmental endpoints relevant to the EDSTP, the test could be a useful component in screening and testing. Although mysids would provide some indication of endocrine effects for arthropods, it is unlikely to identify vertebrate effects.

The other common invertebrate bioassay, one using the water flea, Daphnia, is used internationally (OECD 202). It incorporates life cycle assessment and reproductive and developmental endpoints, albeit applied quite differently in this group of animals. Reproduction is usually parthenogenic in the laboratory in these animals, limiting the applicability to endpoints identified in this report. The particular aspect of this system is that the Daphnia is sensitive to estrogenic compounds (Baldwin et al., 1995; Baldwin et al., 1997; Shurin and Dodson, 1997), and possesses receptors for testosterone, making the system sensitive to another vertebrate hormone. Again, this bioassay would have to be adapted for the endpoints and process of interest in the EDSTP as a protocol for including invertebrate species in the endpoints addressed by the EDSTP screening and testing battery. Other invertebrates, such as molluscs, crayfishes, and echinoderms, do have EA systems, but again relevant standardized tests for evaluating the consequences of interfering with these systems are not currently available. We simply do not know whether one (mysid) or two (mysid and daphnia) Tier 2 tests will provide sufficiently valid information for other invertebrate groups not tested. This is a source of uncertainty, potentially leading to Type II errors of unknown magnitude.

4. Amphibian Development and Reproduction

A definitive amphibian test, which exposes larvae through metamorphosis and reproduction, is important to evaluate the consequences of endocrine disruption in a poikilothermic oviparous vertebrate distinct from fishes. A rich literature on metamorphosis, growth, and reproduction exists for frogs and promising methods are being developed. No established method has been identified which is suitably comprehensive to stand as a Tier 2 test at this time. The EDSTAC feels a test to address this taxonomic group and set of endpoints is needed in T2T and should be given a high priority for development and standardization.

VII. Validation, Standardization, Methods Development, and Research

A. Concept of Assay Validation and Standardization

As stated earlier, the role of standardization and validation is to provide sufficient data to allow informed decisions about the relative merits of the recommended T1S battery component assays and alternative assays (based on sensitivity, specificity, technical complexity, inter- and intra-laboratory variability, time, and cost).

Validation is the scientific process by which the reliability and relevance of an assay method are evaluated for the purpose of supporting a specific use (ICCVAM, 1996). Relevance refers to the ability of the assay to measure the biological effect of interest. Measures of relevance can include sensitivity (the ability to detect positive effects), specificity (the ability to give negative results for chemicals that do not cause the effect of interest), statistically derived correlation coefficients, and determination of the mechanism of the assay response with the toxic effects of interest. Reliability is an objective measure of a method's intra- and inter-laboratory reproducibility. The process of validation includes standardization, that is, definition of conditions under which the assay is run (species, strain, culture medium, dosing regimen, etc.). Standardization is critical to ensure reliability, that is, valid results from time to time and between laboratories. Even in those instances where there is currently some degree of *de facto* acceptance of a given screening method as valid, there is a need for such standardization.

B. Statutory Need for Validation

The Food Quality Protection Act (FQPA) requires EPA "to develop a screening program, using appropriate validated test systems and other scientifically relevant information, to determine whether certain substances may have an effect in humans that is similar to an effect produced by a naturally occurring estrogen, or other such endocrine effect as the Administrator shall designate" by August 1998 with implementation of the peer reviewed program by August 1999. In requiring the use of validated test systems the FQPA is merely mandating good science. There are numerous reasons for using only validated assays. These include: having confidence that the assay is detecting the effect it purports to be detecting, that the results of the assay are reproducible and comparable from laboratory to laboratory, and that the results permit a comparison of the toxicity of various chemicals. These factors are important in being able to interpret results to establish a relative priority among chemicals for progressing from screening to testing and ultimately to perform a hazard and risk assessment.

C. Addressing the Validation Issue

The assays and tests recommended by the EDSTAC vary considerably in terms of their degree of development and validation. EPA (Dr. Lynn Goldman, April 24, 1997, letter to EDSTAC) recognized that few screening assays have actually met the "gold standard" of validation and that other assays have been accepted on the basis of peer review and general use without formal validation. Because the number of validated off-the-shelf assays is so limited, EDSTAC was asked by EPA to extend its consideration to all existing assays.

Thus, although formal validation was not a prerequisite for assay selection by the EDSTAC, the ability or potential of an assay to be validated must be considered because assays must be capable of passing the "validation test" before the screening program is fully implemented. The following are a list of factors that EPA might consider in estimating the likelihood that a candidate assay will actually survive the validation process. If possible, it would be useful to devise a quantitative or semi-quantitative scale for expressing these characteristics so that test methods reviewed by various people could more easily be compared.

1. Number of independent, peer-reviewed publications reporting results from the assay.
2. Similarity of results from independent publications performing the assay.
3. Number of independent laboratories publishing results from the assay.
4. Consistency of the methods used across laboratories.
5. Consistency of results of the assay between labs (to the extent results are available for the same chemicals).
6. Known variability of the assay within single laboratories (may not preclude use as a valuable research tool, but may have important implications for use as a widespread screening tool).
7. Age of the assay - is it an "old standby" or a "new kid on the block."
8. The extent to which the assay relies upon calibrated equipment and calibrated standards.
9. The extent to which the assay depends upon the skill of the technician (a professional opinion from those who know about the assay).
10. The extent to which the assay utilizes internal controls or standards.
11. Use of the assay to develop clinically effective drugs (although not of itself proof of validity, success speaks well for itself).
12. Number of examples of false positives/false negatives from the assay (although we have a clear charge to eliminate false negatives at the screening stage, it is nonetheless important to consider the overall performance of an assay in order to estimate its likelihood of being validatable).
13. Any known species- or strain-specific sensitivities of the assay. For *in vitro* assays, any knowledge of critical sensitivity to cell characteristics such as passage number, plating density, doubling time, etc. or other specific sensitivities of the assay, such as receptor number, transfection technique, serum requirements, media composition, etc. (trying to get another angle on how finicky the assay is).

Information related to validation status of the assays is summarized in Chapter Five, Section VII, E, and was among the factors considered in deciding among assays. These same considerations regarding validation and standardization apply to T2T.

D. Validation and Standardization Process

While not all assays would necessarily need to be validated since they may have *de facto* acceptance as being valid in the scientific community due to their long history of use and

performance, others have little or no data that would allow judgments to be made regarding their validity.

The following is a description of elements of a such a validation process for endocrine disruptor screening assays.

1. Characterize Reference Substances/Vehicles.
    - A reference substance for each hormone endpoint (minimum)
    - Reference substances for each hormone endpoint:
        - positive control substances
        - negative control substances
    - Natural and man-made substances
    - Composition and purity defined (e.g., GC/MS)
    - Stability verified
    - Coded
    - Centralized distribution (using same batch number, lot number, etc.)

2. Develop a Standard Protocol for Each Assay Method.
    - Assay system to be used (species, strain, sex, age, cell line, clone, gene construct, etc.)
    - Dose levels/exposure concentrations
    - General criteria for selection
    - MTD (whole animal systems; e.g., mortality, decreased body weight, etc.)
    - Cell viability (*in vitro* systems): 2 methods
    - Solubility limitations
    - Specific for each reference substance
    - Dose/exposure regimen
    - Number of doses
    - Duration
    - Guidance of mixtures
    - Concentrations of reference substances
    - Interpretation of data
    - Route of exposure (whole animal systems)
    - Description of endpoint(s) to be measured
    - Materials (equipment, media, vehicles, etc.)
    - Time(s) of measurements
    - Criteria for positive/negative response
    - Statistical methods to used
    - Number of replications required (depending on the study design)

3. Define Specialized Skills and Equipment Required for Each Assay Method

4. Conduct in a Variety of Laboratories.

5. Compile and Evaluate Data
    - Expert scientific oversight group
    - Inter-laboratory/intra-laboratory variability
    - Positive versus negative response
    - Relative potency (in comparison to reference substance)
    - Viability/maintenance of assay system (including passage number and growth curves)
    - Sensitivity of assay system (e.g., minimal effective dose/concentration)
    - Specificity of assay system (positive versus negative controls)

E. Levels of Effort Necessary to Validate the Recommended Screens and Tests

The EDSTAC believes validation and standardization of the recommended screens and tests are essential for implementation of the EDSTP. The EDSTAC also believes the validation and standardization program is of highest priority, and recommends that it proceed on an accelerated schedule. As indicated earlier, the EDSTAC recommends that the validation and standardization program be consistent with the principles articulated by the national (ICCVAM, 1996; Zeiger, 1998) and international (OECD, 1996) alternative methods validation groups. As mentioned throughout the chapter, each screen and test recommended for T1S or T2T needs some level of validation, standardization, methods development, or further research before being accepted as a regulatory toxicity screen or test for inclusion in the EDSTP. The level of effort needed to fully standardize and validate may be different for each individual screen or test (including all recommended endpoints). The effort required for each screen or test will be defined by a variety of criteria including: period of time in use, existing level of general acceptance in the endocrine toxicology field, and existing understanding of relevancy and reliability. Regardless of the effort required, EDSTAC believes all the screens and tests recommended for T1S and T2T must be fully validated and standardized before being included in the EDSTP. The EDSTAC recommends that as individual screens and tests are validated and standardized, they can be utilized in the EDSTP without waiting for all screens and tests in the batteries to be validated. The EDSTAC further recommends that a multi-stakeholder process, involving government, industry, and academics, be utilized in validating and standardizing the T1S and T2T batteries (see Section VII, G, of this chapter, for more explanation).

As mentioned earlier, the screens and tests being recommended by the EDSTAC vary considerably in terms of the effort necessary to be fully validated and standardized. Each screen or test lies somewhere along a spectrum of validation from already fully validated and standardized to needing further research to determine their utility in the EDSTP. Outlined in the following few paragraphs, in some detail, is the spectrum of possible effort needed to validate and standardize the screens and tests. In addition, the Screening and Testing Work Group attempted to categorize the levels of validation needed of the recommended screens and tests and their preliminary efforts are found in Appendix R. It is possible that the status of these screens and tests may have changed as work has progressed in the interim. The EDSTAC recommends that EPA update this categorization scheme as part of their validation and standardization program.

The recommended screens and tests (including all endpoints) will have to meet all the criteria of relevance and reliability for use in regulatory toxicity screening or testing for EAT in order to be considered fully validated and standardized (ICCVAM, 1996; Zeiger, 1998). As screens and tests become fully validated and standardized, they will warrant inclusion in the EDSTP according to their specific and appropriate use. None of the screens, new tests, or enhancements to existing test guidelines included in T1S or T2T completely fulfill these criteria to date.

Some of the recommended screens and tests have been in use for a sufficient period of time and have therefore gained sufficient general acceptance within the field of endocrine toxicology to be considered *de facto* validated (reliable *and* relevant). These assays measure relevant endpoints, are responsive to endocrine active compounds with a high degree of specificity, are sufficiently sensitive to identify all known active agents, and can reasonably be expected to give reproducible results from laboratory to laboratory, assuming a general level of competence and expertise. Nonetheless, variations in protocols for these screens and tests can produce disparate results. Therefore, before the recommended screens and tests are utilized in the EDSTP, a standard performance guideline should be developed that produces an acceptable level of consistency in results for each one.

Some of the recommended screens and tests have sufficiently broad use to be generally considered relevant OR reliable to either screening for endocrine activity (Tier 1) or to testing for adverse endocrine-mediated effects (Tier 2). These screens and tests cannot, however, be generally considered to be both relevant *and* reliable. The level of performance that can be expected of these screens and tests with respect to identifying endocrine active agents or endocrine disruptive effects of chemicals must be clarified. Therefore, these assays should undergo further but focused validation and standardization to define their relevance and reliability for the task of endocrine disruptor screening or testing. The validation required may be focused to answer specific questions about relevance and to provide information regarding specificity and sensitivity.

Some of the recommended screens and tests may have relevance to the task of either screening for endocrine activity or testing for endocrine disruptive effects, but their performance in identifying endocrine active agents or endocrine disruptive effects has seen only limited testing. Questions as to whether these assays measure endpoints that are relevant to endocrine activity or endocrine disruptive effects, whether these assays respond with specificity and sensitivity to known endocrine active agents, or whether they identify endocrine disruptive effects cannot be addressed with information currently available. In addition, questions regarding the specific protocols and conditions under which the assays should be conducted must be answered before relevance and reliability can be assessed. Nonetheless, the EDSTAC feels these assays would have sufficient utility, if further developed and validated, to enhance or augment the screening and testing program. Therefore, the EDSTAC recommends that resources be made available to pursue methods development and validation and standardization of these assays.

F. Screens and Tests Recommended for Further Research

As discussed in Section III, A, 5, the EDSTAC recognizes the importance of evaluating postnatal consequences of *in utero* and *in ovo* exposures to chemicals with EAT activities. Such an assay has not been incorporated into T1S due, in part, to the lack of an appropriate, short-term, cost-effective assay. The EDSTAC, however, does recommend that EPA take the lead in designing, standardizing, and validating such an assay. The EDSTAC also identified other screens and tests which, if available, could have an important utility in the screening and testing program, and recommends that research be conducted to determine whether such assays can be developed, and if so, what purpose the assays could fulfill within the endocrine disruptor screening and testing program. The rationale for including each of the following assays and tests is found below.

- *in utero* developmental screening assay
- *in ovo* developmental screening assay
- avian androgenicity screening assay
- invertebrate screening assays
- amphibian development and reproduction test (*Xenopus*)
- reptilian reproduction test

In Utero Developmental Screening Assay: The rationale for developing an *in utero* developmental screening assay was discussed earlier; however, the EDSTAC did consider one possible study design which is summarized here (and further elaborated upon in Appendix O). In the study design, pregnant rats would be exposed to the tested chemical from the start of the embryonic period through weaning of the offspring three weeks after birth. Types of EAT-sensitive endpoints that could be measured in this assay might include apparent sex ratio (based on anogenital distance), numbers of offspring per litter, anogenital distance, retention of nipples in males, precocious puberty in females, uterine and ovarian weight and various uterine histological and biochemical parameters in females, and reproductive tract anomalies in both sexes. Blood samples could be analyzed for estradiol, testosterone, T4, and TSH. Myelin basic protein in brains could be measured as an indicator of thyroid hormone activity.

How much effort such an assay would entail or what its cost might be is not currently known. There may be additional protocols that could be predictive of EAT in developing systems and which are more amenable to screening applications than the protocol included and therefore the EDSTAC encourages development of other assays to address this issue.

In Ovo Developmental Screening Assay: A major route of excretion of lipophilic contaminants for female birds is into the yolk of their eggs; therefore, their embryos can have high levels of exposure from the earliest stages of development. In addition, the endocrine control of sexual and reproductive development is fundamentally different in birds than in mammals. Hence, a short-term screening assay for chemical substances or mixtures that alter avian development is highly desirable. There is a moderate amount of research on the effects of environmental contaminants injected into bird eggs that could be the basis for developing such an assay. The more general rationale for developing an *in ovo* developmental screening assay, as with an *in utero* one, is found earlier in the chapter.

Avian Androgenicity Screening Assay: This assay would be useful in the T1S battery to improve and extend our assessment of chemical substances or mixtures for androgenic and antiandrogenic activity in birds. Development of this screen will become important if data from T2T point to differences in the actions of chemical substances or mixtures in birds versus mammals.

Reptilian Reproduction Test: Several distinctive features of reptilian reproduction (e.g., ovoviviparity, temperature-dependent sex determination), and a generally long life span that allows high body burdens of environmental contaminants to accumulate in reptiles, underscore the importance of developing a practical reproductive test in this class of ecologically important vertebrates.

G. Stakeholder Involvement in the Validation Program

The EDSTAC recommends that a multi-stakeholder process involving government, industry, and academics be utilized in standardizing and validating the T1S and T2T batteries. One key step in instituting a validation program for T1S assays is the identification of a set of "standard test substances" for the individual assays as well as for the overall T1S battery. To the extent possible, the standard test substances will be chosen according to the following criteria:

1. known EAT positives which act via receptor binding;
2. known EAT positives that do not appear to act via receptor binding (i.e., via some other mechanism such as alterations of hormone synthesis, degradation, transport, etc.);
3. known EAT negatives (i.e., substances known not to have hormonal activity);
4. known EAT positives which are active as the parent compound;
5. known EAT positives which require metabolic activation;
6. substances that cover a wide range of EAT potencies;
7. substances with a wide range of physical properties (e.g., pH, reactivity, volatility, etc.); and
8. substances with extensive *in vivo* databases with *in vivo* effects that have been well documented.

It may not be possible to satisfy every one of the above criteria (e.g., there are currently no known examples of environmental thyroid or androgen receptor agonists), but every standard test substance selected should meet at least one of the criteria.

In addition, careful definition of the expected use of the set of chemicals is necessary to avoid inappropriate use. Such a set of chemicals, developed with the already mentioned criteria in mind, would be used in the validation program to assist in defining their relevance and reliability for the task of endocrine disruptor screening, (i.e., to identify whether a specific chemical substance or mixture has endocrine activity, or can be placed in the "hold box").

Further, as was also stated earlier, it is critical to acknowledge state of the science in this area is evolving rapidly, and assays currently being developed, or ones developed in the future, may offer

distinct advantages over some included in the current options. As they are developed, validated, and standardized, the use of these new assays for screening is strongly encouraged.

H. Preliminary Cost Estimates for the T1S and T2T Batteries

An EDSTAC member completed a survey of contract labs in an effort to estimate the cost of the recommended T1S and T2T batteries. The results of this survey are summarized in Tables 5.6 and 5.7 and Appendix S. A detailed description of the methodology used to conduct the survey is included in the EDSTAC Docket (#OPPTS-42189, TSCA Public Docket Office, U.S. EPA). The EDSTAC includes these estimates in its final report, but acknowledges they are both preliminary and uncertain given the inherent uncertainties regarding the outcome of the validation and standardization process. The EDSTAC also notes that these cost estimates should be viewed in the context of the near- and long-term public health and environmental protection benefits to society.

VIII. Compilation of Chapter Five Recommendations

A. Tier 1 Screening

1. The EDSTAC recommends that any T1S battery designed to detect endocrine disruptors should meet five criteria. The battery should:

- maximize sensitivity to minimize false negatives while permitting an acceptable level of false positives;
- include a range of organisms representing known or anticipated differences in metabolic activity;
- detect all known modes of action for endocrine endpoints of concern;
- include a sufficient range of taxonomic groups among the test organisms; and
- incorporate sufficient diversity and complementarity among the endpoints and assays to reach conclusions based on weight of evidence considerations.

2. The EDSTAC recommends the following assays for inclusion in the T1S battery:

In Vitro

1. Estrogen Receptor (ER)Binding/Transcriptional Activation Assay;
2. Androgen Receptor (AR) Binding/Transcriptional Activation Assay; and
3. Steroidogenesis Assay With Minced Testis.

In Vivo

1. Rodent 3-day Uterotrophic Assay (subcutaneous);
2. Rodent 20-day Pubertal Female Assay With Thyroid;
3. Rodent 5-7 day Hershberger Assay;
4. Frog Metamorphosis Assay; and
5. Fish Gonadal Recrudescence Assay.

3. The EDSTAC identified the following four assays as possible alternatives to some components of the proposed battery and recommends that they also be standardized and validated:

In Vitro

1. Placental Aromatase Assay

In Vivo

1. Modified Rodent 3-day Uterotrophic Assay (intraperitoneal);
2. Rodent 14-day Intact Adult Male Assay With Thyroid; and
3. Rodent 20-day Thyroid/Pubertal Male Assay.

Combinations of the alternative assays, if validated and found to be functionally equivalent, could potentially replace three of the component assays in the proposed T1S battery (*in vitro* steroidogenesis assay with testis, 20-day pubertal female assay, and 5-7 day Hershberger

assay) thereby possibly reducing the overall time, cost, and complexity while maintaining equivalent performances of the overall T1S battery. The EDSTAC recognizes that the state-of-the-science in this area is evolving quickly and strongly encourages the use of new or improved assays for screening as they become available.

4. The EDSTAC recommends that validation/standardization studies be conducted on all assays in the proposed battery as well as the alternatives.

5. The EDSTAC agrees that EPA should take affirmative steps, in collaboration with industry and other interested parties, to attempt to develop the protocol for a full life cycle (i.e., with embryonic exposure and evaluation of the adult offspring) developmental exposure screening assay that can be subjected to validation and standardization. The EDSTAC further recommends that, if such an assay were identified, validated, and standardized, the decision on whether it should be included in the T1S battery should include an evaluation of its potential to replace one or more of the recommended T1S assays and its overall impact on the cost effectiveness of the T1S battery.

6. The EDSTAC recommends that all T1S *in vitro* assays involve multiple dose levels, whether performed by HTPS or bench level methods, so a dose-response curve and assessment of relative potencies can be developed. Subject to the results of the validation process, the EDSTAC recommends using one or more dose levels in the performance of the *in vivo* assays.

7. For assessing receptor binding *in vitro*, the EDSTAC is recommending that both the cell-free receptor binding assays and the transcriptional activation assays for ER and AR be incorporated into the T1S battery, and be subjected to validation and standardization.

8. As noted in Chapter Four, the EDSTAC recommends the use of a high throughput pre-screen (HTPS) for toxicants operating though the ER, AR, and TR using stably transfected cell lines with and without metabolic activation, if available. Substances which have not been assessed in the HTPS should be subject to assays for detection of ER and AR activity performed at the bench. Two types of assays are considered acceptable: cell free receptor binding and transcriptional activation in transfected cells. The latter is preferred. Assays must meet the following characteristics:

    - evaluate binding to EAT receptors;
    - evaluate binding with and without metabolic capability;
    - distinguish between agonist and antagonist potential; and
    - yield dose responses to establish relative potency.

9. The EDSTAC is recommending evaluation of antithyroid effects in animals in the longer term rodent screen (either 14-day or 20-day exposure). Although it is not known whether exposure to xenobiotics for greater than 14 days is required to significantly affect circulating levels of T4, TSH, or thyroid histopathology, the EDSTAC believes these longer periods may be required. The effects of duration of chemical substance and mixture exposure must be quickly evaluated in the validation phase.

B. Principles for Evaluating Tier 1 and Tier 2 Results

10. The EDSTAC recommends that a "weight-of-evidence" approach be used in evaluating T1S and T2T results and has developed general criteria for applying "weight-of-evidence" to ensure that decisions are transparent and predictable.

C. Tier 2 Testing

11. The EDSTAC recommends that the following tests be included in the Tier 2 battery:

    Mammalian tests
    - Two-generation reproductive toxicity study, or
    - An alternative less comprehensive test:
        1. Alternative mammalian reproductive; and
        2. One-generation test.

    Non-mammalian multigeneraton tests
    - Avian reproduction;
    - Fish life cycle;
    - Mysid life cycle; and
    - Amphibian development and reproduction.

12. The EDSTAC recommends that the "default" action, in absence of any prior information, be to perform all tests in the T2T battery with all endpoints. Further, the EDSTAC recommends that the choice of whether Tier 2 tests will be conducted on all five of the recommend taxa, or a more limited subset of the five taxa, should be based on the physico-chemical characteristics and environmental release and exposure information of the chemical substance or mixture to be tested, together with biological data from T1S. The results of T1S or other information may also allow tailoring of T2T such as the inclusion or deletion of certain endpoints (e.g., thyroid effects) or use of alternative tests.

13. The EDSTAC believes that a project is required to resolve the underlying uncertainties and controversy about issues related to low dose selection and the identification of no-observed-adverse-effect-levels (NOAEL). Further, the EDSTAC recommends that a collaborative group involving government, industry, and appropriate individuals in academia design the study protocols, be kept abreast of the conduct of the studies, evaluate results, and develop overall conclusions and recommendations.

14. The EDSTAC recommends that information used to select doses in the performance of Tier 2 tests include:

- existing information such as that available during priority setting including the results or the HTPS;
- results from T1S;
- results from other assays or tests; and
- results from range finding studies.

15. The EDSTAC recommends including thyroid-sensitive endpoints in T2T and that dosing in mammalian tests include fetal and lactational exposure.

D. Validation of the Screening and Testing Batteries

16. The EDSTAC believes the validation and standardization program is of highest priority, and recommends that it proceed on an accelerated schedule. The EDSTAC further recommends that the validation and standardization program be consistent with the principles articulated by the national (ICCVAM, 1996; Zeiger, 1998) and international (OECD, 1996) alternative methods validation groups.

17. The EDSTAC recommends that, as individual screens and tests are validated and standardized, they can be utilized in the EDSTP without waiting for all screens and tests in the batteries to be validated.

18. The EDSTAC recommends that a multi-stakeholder process, involving government, industry, and academics, be utilized in validating and standardizing the T1S and T2T batteries.

19. The EDSTAC identified other screens and tests which, if available, could have an important utility in the screening and testing program, and recommends that research be conducted to determine whether such assays can be developed, and if so, what purpose the assays could fulfill within the endocrine disruptor screening and testing program.

IX. Literature Cited

Anderson, J., J.H. Clark, and E.J. Peck, Jr., "Oestrogen and nuclear binding sites. Determination of specific sites by (3 H) oestradiol exchange," *Biochem J*, 126, 1972, pp. 561-7.

Anderson, S., S. Pearce, P. Fail, B. McTaggert, R. Tyl, and L.E. Gray Jr., "Validation of the alternative reproductive test protocol (ART) to assess toxicity of methoxychlor in rats," *The Toxicologist*, 15, 1995, pp. 164.

Anderson, S., S. Pearce, P. Fail, B. McTaggert, R. Tyl, and L.E. Gray Jr., "Testicular and adrenal response in adult Long-Evans Hooded rats after antiandrogenic vinclozolin exposure," *Journal Andrology*,16, 1995, pp. 43.

Badia, E., M.J. Duchesne, S. Fournier-Bidoz, A.E. Simar-Blanchet, B. Terouanne, J.C. Nicolas, and M. Pons, " Hydroxytamoxifen induces a rapid and irreversible inactivation of an estrogenic response in an MCF-7-derived cell line," *Cancer Research*, 54(22), 1994, pp. 5860-5866.

Baldwin, W.S., D.L. Milam, and G.A. LeBlanc, "Physiological and biochemical perturbations in *Daphnia magna* following exposure to the model environmental estrogen diethylstilbestrol," *Environ.Toxicol.Chem.*,14, 1995, pp. 945-952.

Baldwin, W.S., S.E. Graham, D. Shea, and G.A. LeBlanc, "Metabolic androgenization of female *Daphnia magna* by the xenoestrogen 4-nonylphenol," *Environ.Toxicol.Chem.*, 16, 1997, pp. 1905-1911.

Baxter, W.L., R.L. Linder, and R.B. Dahlgren, "Dieldrin Effects in Two Generations of Penned Hen Pheasants" *J. Wildl. Mgmt.*, 33(1), 1969, pp. 96-102.

Bellabarba, D., S. Belisle, N. Gallo-Payet, and J.G. Lehoux, "Mechanism of Action of Thyroid Hormones During Chick Embryogenesis," *Amer. Zool.*,28, 1988, pp. 389-399.

Benoit, D.A., "*User's Guide for Conducting Life-Cycle chronic Toxicity Tests with Fathead Minnows (Pimephales promelas),*" U. S. Environmental Protection Agency, Office of Research and Development, 1981, EPA-600/8-81-011.

Bhat, N.B., G. Shanker, and R.A. Peringer, "Investigations on myelination *in vitro*: regulation of 2',3'-cyclic nucleotide 3'-phosphohydrolasc by thyroid hormone in cultures of dissociated brain cells from embroyonic," *J. Neurochem.*, 37, 1981, pp. 695-701.

Bhat, N.R., L.L. Sarlieve, G.S. Rao et al., "Investigations on myclination *in vitro*: regulation by thyroid hormone in cultures of dissociated brain cells from embryonic micc.," *J. Biol Chem*., 254, 1978, pp. 9342-9344.

Bjerke, D., and R. Peterson, "Reproductive toxicity of 2,3,7,8 tetrachlorodibenzo-p-dioxin in male rats: Different effects of *in utero* versus lactational exposure," *Toxicology and Applied Pharmacology*, 127, 1994, pp. 241-249.

Borst, D.W., H. Laufer, M. Landau, E.S. Chang, W.A. Hertz, F.C. Baker, and D. A. Schooley, "Methyl farnesoate and is role in crustacean reproduction and development," *Insect Biochem*,17, 1987, pp. 1123-1127.

Bradley, D. J., W.S. Young, and I.C. Weinberger, "Differential expression of alpha and beta thyroid hormone receptor genes in rat brain and pituitary," *Proc Natl Acad Sci USA*, 86, 1989, pp. 7250-7254.

Brown, D.D., Z. Wang, J.D. Furlow, A. Kanamori, R.A. Schwartzman, B.F. Remo, and A. Pinder, "The thyroid hormone-induced tail resorption program during Xenopus laevis metamorphosis," *Proc Natl Acad Sci USA*, 93, 1996, pp. 1924-1929.

Brown, D.D., Z. Wang, A. Kanamori, B. Eliceiri, J.D. Furlow, and R. Schwartzman, "Amphibian metamorphosis: a complex program of gene expression changes controlled by the thyroid hormone," *Recent Prog Horm Res*, 50, 1995, pp. 309-315.

Carr, F.E., C.U. Fisher, H.G. Fein, and R.C. Smallridge, "Thyrotropin-releasing hormone stimulates c-jun and c-fos messenger ribonucleic acid levels: implications for calcium mobilization and protein kinase-C activation," *Endocrinology*, 133, 1993, pp. 1700-1707.

Cavailles, V., M.Garcia, H. Rochefort, "Regulation of Cathepsin-D and pS2 Gene Expression by Growth Factors in MCF7 Human Breast Cancer Cells," *Mol Endocrinology*, 3, 1989, pp. 552-558.

Chamberlain, J.G., "Thalidomide and lack of teratogenesis in Long-Evans rats," *Teratology*, 19(1), 1979, pp.129.

Chu, K.H., C.K. Wong, and K.C. Chiu, "Effects of Insect Growth Regulator (S)-Methoprene On Survival And Reproduction Of The Freshwater Cladoceran Moina Macrocopa," *Environ Pollution*, 96(2), 1997, pp.173-178.

Cook, J.C., et al., "Development of a tier I screening battery for detecting endocrine-active compounds," *Regul. Toxicol. Pharmacol.,* 26, 1997, pp. 60-68.

Connor, K., J. Howell, I. Chen, K. Berhane, C. Sciarretta, S. Safe, T. and Zacharewski, "Failure of chloro-S-triazine-derived compounds to induce estrogen receptor-mediated responses *In Vivo* and *In Vitro," Fund Appl Toxicol.,*30, 1996, pp. 93-101.

Cruickhank, J.J., and J.S. Sim, "Morphometric and Radiographic Characteristics Of Tibia Bone of Broiler Chickens with Twisted Leg Disorders," *Avian Diseases,* 30(4), 1986, pp. 699-708.

Dahlgren, R.B., and R.L. Linder, "Effects Of Polychlorinated Biphenyls On Pheasant Reproduction, Behavior and Survival," *J. Wildl. Mgmt.,*35(2), 1971, pp. 315-319.

Davis M.D., W.B. Butler, and S.C. Brooks "Induction of tissue plasminogen activator mRNA and activity by structurally altered estrogens," *J Steroid Biochem Mol Biol.,* 52, 1995, pp. 421-430.

Denver, R.J., S. Pavgi, and Y.B. Shi, "Thyroid hormone-dependent gene expression program for Xenopus neural development," *J Biol Chem*, 272, 1997, pp.8179-8188.

DeVito, M., K. Crofton, and S. McMaster, "Screening Methods for Chemicals That Alter Thyroid Hormone Action, Function and Homeostasis," June 23-25, 1997, Workshop Report EPA/600/R-98/057.

Dodds E., W. Lawson, and R. Noble, "Biological effects of the synthetic ostrogenic substance 4:4'-dihydroxy-:B-diethystilbene." *Lancet,* 1997, 139: 627.

Dussault, J.H., and J. Ruel, "Thyroid hormones and brain development," *Annu Rev Physiol*, 49, 1987, pp. 321-334.

Edgren, R., "Issues in animal pharmacology," *Pharmacology of the contraceptive steroids,* Ed J. Goldzieher, Raven Press, Ltd., New York, 1984

Emlen, Jr., J.T. "Determinants of Cliff Edge and Escape Responses In Herring Gull Chicks in Nature," B*ehaviour,* 22, 1963, pp. 1-15.

Escobar, G., M.J. Obregon, and F. Rey, "Contribution of maternal thyroxine to fetal thyroxine pools in normal rats near term," *Endocrinology*, 126, 1990, pp. 2765-2767.

Escobar, G., M.J. Obregon, and F. Rey, "Fetal and maternal thyroid hormones," *Hormone Res,* 26, 1987, pp. 12-27.

Escobar, G., M.J. Obregon, and F. Rey, "Transfer of thyroid hormones from the mother of the fetus," In Delang, F., D.A. Fisher, and D. Glinoer (Eds.), *Research in Congenital Hypothyroidism Plenum Press*: New York, NY, 1988, pp. 15-28.

Farsotti, A., T. Mitsubashi, B. Desvergne, J. Robbins, and V.M. Nikodem, "Molecular basis of thyroid hormone regulation of myelin basis protein gene expression in rodent brain," *J. Boil. Chem.,* 266, 1991, pp.23226-23232.

Fawcett, D.W., *Bloom and Fawcett: Textbook of Histology*, 11th ed. Philadelphia: W.B. Saunders Company, 1986, pp. 973.

Figueiredo. B.C., G. Almazan, Y. Ma, and W. Tetzlaff, "Gene expression in the developing cerebellum during perinatal hypo- and hyperthyroidism," *Brain Res Mol Brain Res,* 17, 1993, pp. 258-268.

Fleming, W.J., G.H. Heinz, and C.A. Schuler, "Lethal and Behavioral Effects of Chlordimeform in Bobwhite," *Toxicology,* 36, 1985, pp. 37-47.

Fleming, W. J., G.H. Heinz, J.C. Franson, and B.A. Rattner, "Toxicity of Abate 4E (Temephos) in Mallard Ducklings and the Influence of Cold," *Environ.Toxicol. Chem*., 4, 1985, pp.193-199.

Forman, B.M., and H.H. Samuels, "Interactions among a subfamily of nuclear hormone receptors: the regulatory zipper model," *Molec Endocrinol*, 4, 1990, pp. 1293-1301.

Fort, D. J., and E.I. Stover, "Development of Short-Term, Whole-Embryo Assays to Evaluate Detrimental Effects on Amphibian Limb Development and Metamorphisis," *Xenopus laevis, Environmental Toxicology and Risk Assessment: Modeling and Risk Assessment* (Sixth Volume), ASTM STP 1317, Dwyer, F.J., T.R. Doane, and M.I. Hinman (Eds.), American Society for Testing and Materials, Philadelphia, PA, 1997, pp. 376-390.

Fox, G.A., "Eggshell Quality: It's Ecological and Physiological Significance In a DDE-Contaminanted Common Tern Colony," *Wilson Bull*. 88(3), 1976, pp. 459-477.

Fox, G.A., A.P. Gilman, D.B. Peakall, and F.W. Anderka, "Behavioural Abnormalities of Nesting Lake Ontario Herring Gulls," *J. Wildl. Mgmt*., 42, 1978, pp. 477-483.

Fraser, F.C., "Thalidomide retrospective: what did we learn?" *Teratology*, 38(3), 1988, pp.201-202.

Freeman, B.M., and M.A. Vince, "Development of the Avian Embryo," *A Behavioral and Physiological Study*, John Wiley and Sons, New York, NY, 1974, pp. 362.

Franklin, J.A., D.F. Wood, N.J. Balfour, D.B. Ramsden, K. Docherty, W.W. Chin, and M.C. Sheppard, "Effect of hypothyroidism and thyroid hormone replacement *in vivo* on pituitary cytoplasmic concentrations of thyrotropin-ß and alpha-subunit messenger ribonucleic acids," *Endocrinol*., 120, 1987, pp. 2279-2288.

Furlow, J.D., D.L. Berry, Z. Wang, and D.D. Brown, "A set of novel tadpole specific genes expressed only in the epidermis are down-regulated by thyroid hormone during Xenopus laevis metamorphosis," *Dev Biol*, 182, 1997, pp. 284-298.

Gagne, D., P. Balaguer, E. Demirpence, C. Chabret, F. Trousse, J.C. Nicolas, and M. Pons, "Stable luciferase transfected cells for studying steroid receptor biological activity," *Journal of Bioluminescence & Chemiluminescence*, 9, 1994, pp. 201-9.

Gaitan, E. and R.C. Cooksey, "General Concepts of Environmental Goitrogenesis, in Gaitan E, (ed)," *Environmental Goitrogenesis*, 1989, pp. 3-14.

Gaitan, E., R.H. Lindsay, and R.C. Cooksey, "Goiter endemias attributed to chemical and bacterial pollution of water supplies," in Gaitan, E., (Ed.), *Environmental Goitrogenesis*, 1989, pp. 207-232.

Gillesby, B., and T. Zacharewski, "Endocrine disrupters: Mechanism of action of promiscuous receptors and strategies for identification and assessment," *17th Annual Society of Environmental Toxicology and Chemistry Abstract 006*, 1996.

Goldey, E.S., L.S. Kehn, C. Lau, G.L. Rehnberg, and K.M. Crofton, "Developmental exposure to polychlorinated biphenyls (Aroclor 1254) reduces circulating thyroid hormone concentrations and causes hearing deficits in rats," *Toxicol Appl Pharmacol*, 135, 1995, pp. 77-88.

Gonzalez, M.D., F. Lopez, and E. Aguilar, "Involvement of prolactin in the onset of puberty in female rats," *Journal of Endocrinology*, 101, 1983, pp. 63-68.

Gorbman, A., W.W. Dickhoff, S.R. Vigna, N.B.Clark, and C.L. Ralph, *Comparative Endocrinology*, 1983.

Gray L.E. Jr. and J. Ostby, "Effects of pesticides and toxic substances on behavioral and morphological reproductive development: endocrine versus nonendocrine mechansisms." *Toxicol Ind Health*, 14, 1998, pp. 159-184.

Gray, L.E. Jr., W.R. Kelce, T. Wiese, et al., "Endocrine screening methods workshop report: detection of estrogenic and androgenic hormonal and antihormonal activity for chemicals that act via receptor or steroidogenic enzyme mechanisms," *Reprod Toxicol*, 11, 1997a, pp. 719-750.

Gray, L.E. Jr., C. Wolf, P. Mann, and J.S. Ostby, "*In utero* exposure to low doses of 2,3,7,8-tetrachlorodibenzo-p-dioxin alters reproductive development of female Long Evans Hooded Rat offspring," *Toxicology and Applied Pharmacology*, 146, 1997b, pp. 237-244.

Gray, L.E. Jr., J. Ostby, C. Wolf, D. Miller, W. Kelce, C. Gordon, and L. Birnbaum, "Functional developmental toxicity of low doses of 2,3,7,8 tetrachlorodibenzo-p-dioxin and a dioxin-like PCB (169) in Long Evans rats and Syrian hamsters: Reproductive, behavioral and thermoregulatory alterations," *Organohalogen Compounds*, 25, 1995, pp. 33-38.

Gray, L.E. Jr., W.R. Kelce, E. Monosson, J.S. Ostby, and L.S. Birnbaum, "Exposure to TCDD during development permanently alters reproductive function in male LE rats and Hamsters: Reduced ejaculated and epididymal sperm numbers and sex accessory gland weights in offspring with normal androgenic status," *Toxicol and Appl Pharmacol*, 131 (1), 1995, pp.108-118.

Gray, L.E. Jr., J. Ostby, J. Ferrell, G. Rehnberg, R. Linder, R. Cooper, J. Goldman, V. Slott, and J. Laskey, "A dose- response analysis of methoxychlor-induced alterations of reproductive development and function in the rat," *Fundamental and Applied Toxicology*, 12, 1989, pp. 92-108.

Green, W.L, "Antithyroid Compounds," in Braverman, L.E., and R.D. Utiger (Ed.), *Werner and Ingbar's The Thyroid: A Fundamental and Clinical Text*, 1996, pp. 266-267.

Greer, M.A., N. Sato, X. Wang, S.E. Greer, and S. McAdams, "Evidence that the major physiological role of TRH in the hypothalamic paraventricular nuclei may be to regulate the set-point for thyroid hormone negative feedback on the pituitary thyrotroph," *Neuroendocrinol*, 57, 1993, pp. 69-575.

Gustafsson, J.A., "Characteristics and function of a novel estrogen receptor β. Steroid Receptor Superfaniily Symposium," University of Wisconsin, Madison, WI, September 27-29, 1996.

Hannon, W., F. Hill, J. Bernert et al., "Premature thelarche in Puerto Rico: a search for environmental estrogenic contamination," *Arch. Environ. Contam. Toxicol*., 16, 1978, pp. 255-262.

Hansen, D.J., P.R. Parrish, S.C. Schimmel, and L.R. Goodman, "Toxicity Test Using Sheepshead Minnows (*Cyprinodon variegatus*)," Bioassay Procedures for the Ocean Disposal Permit Program, EPA-600/9-78-010, 1978.

Hayes, T.B., "Amphian metamorphosis: An integrative approach," *Amer. Zool*., 37, 1997, pp. 121-123.

Hayes, T.B., "Steroids as potential modulators of thyroid hormone acitivity in anuran metamorphosis," *Amer. Zool*., 37, 1997b, pp. 185-194.

Hershberger, L., E. Shipley, and R. Meyer, "Myotrophic activity of 19-nortestosterone and other steroids determined by modified levator ani muscle method," *Proc Soc Exp Biol Med,* 83, 1953, pp. 175.

Hoffman, D.J., and W.C. Eastin, Jr., "Effects of Malathion, Diazinon, and Parathion on Mallard Embryo Development and Cholinesterase Activity," *Environ. Res*., 15, 1981, pp. 100-107.

Hoffman, D.J., and P.H. Albers, "Evaluation of Potential Embryotoxicity and Teratogenicity of 42 Herbicides, Insecticides and Petroleum Contaminants to Mallard Eggs," *Arch. Environ. Contam. Toxicol*., 13, 1984, pp. 15-27.

Hoffman, D.J., "Embryotoxicity and Teratogenicity of Environmental Contaminants to Bird Eggs," *Rev. Environ. Contam. Toxicol*., 115, 1990, pp.39-89.

Hoffman, D.J., G. J. Smith, and B.A. Rattner, "Biomarkers of Contaminant Exposure in Common Terns and Black-Crowned Night Herons in the Great Lakes," *Environ. Toxicol. Chem.,* 12, 1993, pp. 1095-1103.

Hoffman, D.J., M.J. Melancon, P.N. Klien, C.P. Rice, J.D. Eisemann, R.K. Hines, J.W. Spann, and G.W. Pendleton, "Developmental Toxicity of PCB 126 (3,3',4,4',5-Pentachlorobiphenyl) in Nestling Amercian Kestrels (Falco sparverius)," *Fund. Appl. Toxicol.,* 34, 1996, pp. 188-200.

Hollenberg, A.N., T. Monden, T.R. Flynn, M.E. Boers, O. Cohen, and F.E. Wondisford, "The human thyrotropin-releasing hormone gene is regulated by thyroid hormone through two distinct classes of negative thyroid hormone response elements," *Molecular Endocrinology*, 9, 1995, pp. 540-550.

Hostetter, M., and B. Piacsek, "The effect of prolactin deficiency during sexual maturation in the male rat," *Biology of Reproduction,* 17, 1977, pp. 574-577.

Huhtaniemi, I., A. Amsterdam, and Z. Naor, "Effect of postnatal treatment with a gonadotropin-releasing hormone antagonist on sexual maturation of male rats," *Biology of Reproduction,* 35, 1986, pp. 501-507.

Iniguez, M., A. Rodriguez-Pena, N. Ibarrola, M. Aguilera, G. Escobar, and J. Bernal, "Thyroid hormone regulation of RC3, a brain-specific gene encoding a protein kinase-C substrate," *Endocrinology*, 133, 1993, pp. 467-473.

Jefferies, D.L., and J.L.F. Parslow, "Thyroid Changes in PCB-Dosed Guillemots and Their Indication of One of the Mechanisms of Action For These Materials," *Environ. Pollut.,* 10, 1976, pp. 293-311.

Kanamori, A., and D.D. Brown, "The analysis of complex developmental programmes: amphibian metamorphosis," *Genes Cell,* 1, 1996, pp. 429-435.

Kelce, W.R., E. Monosson, M.P. Gamcsik, S.C. Laws, and L.E. Gray, Jr., "Environmental hormone disruptors: Evidence that vinclozolin developmental toxicity is mediated by antiandrogenic metabolites," *Toxicol Appl Pharmacol,* 126, 1994, pp. 275-285.

Kelce, W.R., C. Stone, S. Laws, L.E. Gray Jr., J. Kemppainen, and E. Wilson, "Persistent DDT metabolite p,p' DDE is a potent androgen receptor antagonist," *Nature*, 375(15), 1995, pp. 581-585.

Kelce, W.R., C. Lambright, L.E. Gray Jr., and K. Roberts, "Vinclozolin and *pp'* DDE alter androgen-dependent gene expression: *in vivo* confirmation of an androgen receptor mediated mechanism." *Toxicol Applied Pharmacology*, 142, 1997, pp. 192-200.

Kliewer, S.A., K. Umesono, D.J. Mangelsdorf, and R.M. Evans, "Retinoid X receptor interacts with nuclear receptors in retinoic acid, thyroid hormone and vitamin D3 signalling," *Nature*, 355, 1992, pp. 446-449.

Klinefelter, G.R., and W.R. Kelce, "A Comparison of Leydig Cell Responsiveness to Hormonal and Nonhormonal Factors *In Vivo and In Vitro*. *The Leydig Cell*, Russel, L., and M. Hardy (Eds.), Cache River Press, 1996.

Koller, K.J., R.S. Wolff, M.K. Warden, and R.T. Zoeller, R. T., "Thyroid hormones regulate levels of thyrotropin-releasing hormone mRNA in the paraventricular nucleus," *Proc Natl Acad Sci USA*, 84, 1987, pp. 7329-7333.

Korach K.S., "Estrogen action in the mouse uterus: characterization of the cytosol and nuclear receptor systems," *Endocrinology* 104, 1979, pp.1324-32. Several Reviews By Katenellenbogen

Korenbrot, C.C., I. Huhtaniemi, and R. Weiner, "Preputial separation ass an external sign of pubertal development in the male rat," *Biology of Reproduction*, 17, 1977, pp. 298-303.

Kubiak, T.J., H.J. Harris, L.M. Smith, T.R. Schwartz, D.L. Stalling, J.A. Trick, L. Sileo, D. Docherty, and T.C. Erdman, "Microcontaminants and Reproductive Impairment of the Forster's Tern on Green Bay, Lake Michigan - 1983," *Arch. Environ. Contam. Toxicol.*, 18, 1989, pp. 706-727.

Laufer, H., D. Borst, F.C. Baker, C. Carrasco, M. Sinkus, C.C. Reuter, L.W. Tsai, and D.A. Schooley, "Identification of a juvenile hormone-like compound in a crustacean," *Science*, 235, 1987, pp. 202-205.

Laufer, H., M. Landau, E. Homola, and D.W. Borst, "Methyl farnesoate: its site of synthesis and regulation of secretion in a juvenile crustacean," *Insect Biochemistry*, 17, 1987b, pp. 1129-1131.

Lazar, M.A., "Thyroid hormone receptors: multiple forms, multiple possibilities," *Endocrine Rev*, 14, 1993, pp. 184-193.

Lazar, M.A., "Thyroid hormone receptors: Update 1994," *Endocrine Reviews Monographs*, 3, 1994, pp. 280-283.

Leatherland, J.F., "Reflections on the thyroidology of fishes: from molecules to humankind," *Guelph Ichthyology Reviews*, 2, 1994, pp. 1-67.

Luke, M., and D. Coffey, "The male sex accessory tissues. Chapter 23.," *The physiology of reproduction*, 2nd edition, Knobil, E. and J. Neill (Eds.), Raven Press, New York, 1994.

Maguire, C.C., and B.A. Williams, "Response of Thermal Stressed Bobwhite to Organophosphorous Exposure," *Environ. Pollut.*, 47, 1987, pp. 25-39.

Mano, H., R. Mori, T. Ozawa, K. Takeyama, Y. Yoshizawa, R. Kojima, Y. Arao, S. Masushige, and S. Kato, "Positive and negative regulation of retinoid X receptor gene expression by thyroid hormone in the rat," *J Biol Chem*, 269, 1994, pp. 1591-1594.

Martin, P.A., "Effects of Carbofuran, Chlorpyrifos, and Deltamethrin on Hatchability, Deformity, Chick Size, and Incubation Time of Japanese Quail (Coturnix japonica) Eggs," *Environ. Toxicol. Chem*., 9, 1990, pp. 529-534.

Martin, P. A., and K.R. Solomon, "Acute Carbofuran Exposure and Cold Stress: Interactive effects in Mallard Ducklings," *Pesticide Biochemistry and Physiology*, 40, 1991, pp. 117-127.

McArthur, M.L.B., G.A. Fox, D.B. Peakall, and B.J.R. Philogene, "Ecological Significance of Behavioral and Hormonal Abnormalities in Bredding Ring Doves Fed an Organochlorine Chemcial Mixture," *Arch Environm. Contam. Toxicol*., 12, 1983, pp. 343-353.

McLachlan, J.A., R.R. Newbold, H.C. Shah, M.D. Hogan, and R.L. Dixon, "Reduced fertility in female mice exposed transplacentally to diethylstilbestrol (DES). *Fertility and Sterility*, 38, 1982, pp. 364-371.

McNabb, F.M.A., "Peripheral Thyroid Hormone Dynamics in Precocial and Altricial Avian Development," *Am Zool.,* 28, 1988, pp. 427-440.

Meyer, T., R. Koop, E. von Angerer, H. Schonenberger, and E. Holler, "A rapid luciferase transfection assay for transcription activation effects and stability control of estrogens in cell cultures," *Journal of Cancer Research & Clinical Oncology*, 120, 1994, pp. 359-64.

Mirell, C.J., M. Yanagisawa, R. Lau, A.E. Pekary, W.W. Chin, and J.M. Hershman, "Influence of thyroidal status on pituitary content of thyrotropin ß- and alpha-subunit, growth hormone, and prolactin messenger ribonucleicacids," *Mol Endocrinol*, 1, 1987, pp.408-412.

Mitsubashi, T., G.E. Tennyson, and V.M. Nikodem, "Alternative splicing generates messages encoding rat c-erbA proteins that do not bind thyroid hormone," *Proc Natl Acad Sci USA*, 85, 1988, pp. 5804-5808.

Moccia, R.D., G.A. Fox, and A. Britton, "A Quantitative Assessment of Thyroid Histopathology Of Herring Gulls (Larus Argentatus) From the Great Lakes and A Hypothesis On the Causal Role Of Environmental Contaminants," *J. Wild. Dis.*, 22(1), 1986, pp. 60-70.

Morley, J.E., "Neuroendocrine control of thyrotropin secretion," *Endocrine Rev*., 2, 1981, pp. 396-436.

Morrissey, R.E., J.D.George, C.J. Price, et al., "The Developmental Toxicity of Bisphenol A in Rats and Mice," *Fund Appl Tox*, 8, 1987, pp. 571-582.

Mortimer, M.R., and D.W. Connell, "Bioconcentration factors and kinetics with a juvenile crab *Portunus pelagicus* (L)," *Australian Journal of Marine and Freshwater Resources*, 44, 1983, pp. 565-576.

Mortimer, M.R., and D.W. Connell, "Critical internal and aqueous lethal concentrations of chlorobenzenes with the crab *Portunus pelagicus* (L)," *Ecotoxicology and Environmental Safety*, 28, 1994, pp. 298-312.

Mortimer, M.R., and D.W. Connell, "Effect of exposure to chlorobenzenes on growth rates of the crab *Portunus pelagicus* (L)," *Environmental Science and Technology*. 29(8), 1995, pp. 1881-1886.

Mortimer, M.R., and D.W. Connell, "A model of the environmental fate of chlorohydorcarbon contaminants associated with Sydney sewage discharges," *Chemosphere*, 30(11), 1995, pp. 2021-2038.

Murray, M., and G. Reidy, "Selectivity in the inhibition of mammalian cytochromes P-450 by chemical agents," *Pharmacological Reviews*, 42, 1990, pp.85-101.

Myant, N.B., "The role of thyroid hormone in the fetal and postnatal development of mammals," *Hormones in Development*, 1971.

Nagel, S.C., F.S. vom Saal, K.A. Thayer, M.G. Dhar, M. Boechler, and W.V. Welshons, "Relative Binding Affinity-Serum Modified Access (RBA-SMA) Assay Predicts the Relative *In Vivo* Bioactivity of the Xenoestrogens Bisphenol A and Octylphenol," *Environmental Health Perspectives*, 105(1), 1997, pp. 70-76.

Nelson, K., T. Takahashi, N. Bossert, D. Walmer, and J. McLachlan, "Epidermal growth factor replaces estrogen in the stimulation of female genital-tract growth and differentiation," 1991.

Nunez, A., M. Berry, J. Imler, P. Chambon, "The 5' flanking region of the pS2 Gene contains a complex enhancer region responsive to oestrogens, epidermal growth factor, a tumor promoter (TPA), the c-Ha-*ras* oncoprotein and the c-*jun* protein," *EMBO J,* 8, 1989, pp. 823-829.

O'Connor, J.C., et al, "An *in vivo* battery for identifying endocrine modulators that are estrogenic or dopamine regulators," *Fund. Appl. Toxicol*. 33, 1996, pp.182-195.

Oppenheimer, J.H., H.L. Schwartz, K.A. Strait, "Thyroid hormone action 1994: the plot thickens," *Eur J Endocrinol*, 130, 1994, pp.15-24.

Parker, A., *Marshall's Physiology of Reproduction*. Parkes, A. (Ed.), Vol. III, Longmans Green and Co. Ltd., London, 1966.

Perez-Castillo, A., J. Bernal, B. Ferreiro, T. Pans, "The early ontogenesis of thyroid hormone receptor in the rat fetus," *Endocrinology*, 117, 1985, pp. 2457-2461.

Petit, F., Y. Valotaire, and F. Pakdel, "Differential functional activities of rainbow trout and human estrogen receptors expressed in the yeast," *Saccharomyces cerevisiae. Eur. J. Biochem.*, 223, 1995, pp. 584-592.

Pilat, M.J., M.S. Hafner, L.G. Kral, and S.C. Brooks, "Differential induction of pS2 and cathepsin D mRNAs by structurally altered estrogens," *Biochemistry*, 32, 1993, pp. 7009-7015.

Pons, M., D. Gagne, J.C. Nicolas, and M. Mehtali, "A new cellular model of response to estrogens: a bioluminescent test to characterize (anti) estrogen molecules," *Biotechniques*, 9, 1990, pp. 450-9.

Porterfield, S.P., "Vulnerability of the developing brain to thyroid abnormalities: environmental insults to the thyroid system," *Environmental Health Perspectives*, 102(2), 1994, pp.125-130.

Porterfield, S.P., and S.A. Stein, "Thyroid hormones and neurological development: update 1994," *Endocrine Rev*, 3, 1994, pp. 357-363.

Porterfield, S.P., and C.E. Hendrich, "The role of thyroid hormones in prenatal neonatal neurological development-current perspectives," *Endocrine Rev*, 14, 1993, pp. 94-106.

Porterfield, S.P., C.E. Hendrich, "Tissue iodothyroidine levels in fetuses of control and hypothyroid rats at 13 and 16 days gestation," *Endocinol*, 131, 1992, pp. 195-106.

Ramely, J., and C. Phares, "Delay of puberty onset in males due to suppression of growth hormone," *Neuroendocrinology*, 36, 1983, pp. 321-329.

Ramirez, V., and C. Sawyer, "Advancement of puberty in the female rat by estrogen," *Endocrinology*, 76, 1964, pp. 1158-1168.

Rattner, B.A., J.M. Becker, and T. Nakatsugawa, "Enhancement of Parathion Toxicity to Quail by Heat and Cold Exposure," *Pesticide Biochemistry and Physiology*, 27, 1987, pp. 330-339.

Rattner, B.A., L. Sileo, and C.G. Scanes, "Hormonal Responses and Tolerance to Cold of Female Quail following Parathion Ingestion," *Pest. Biochem. Physiol*., 18, 1982, pp.132-138.

Raynaud, J.P., C. Bonne, M. Moguilewsky, F.A. Lefebvre, A. Belanger, and F. Labrie, *The Prostate,* 5, 1984, pp. 299-311.

Reel, J., A. Lawton, R. Wolkowski-Tyl, G. Davis, and J. Lamb, "Evaluation of a new reproductive toxicology protocol using diethylstilbestrol (DES) as a positive control compound," *Journal of the American College of Toxicology,* 4, 1985, pp. 147.

Reel, J., J. Lamb, and B. Neal, "Survey and assessment of mammalian estrogen biological assays for hazard characterization," *Fundamental and Applied Toxicology,* 33, 1996.

Rodriguez-Pena, A., N. Ibarrola, M.A. Iniguez, A. Munoz, and J. Bernal, "Neonatal hypothyroidism affects the timely expression of myelin-associated glycoprotein in the rat brain," *J. Clin Invest*., 91, 1993, pp. 812-818.

Russo, I.H., and J. Russo, "Mammary gland neoplasia in long-term rodent studies," *Environmental Health Perspectives*, 104, 1996, pp. 938-967.

Salamon, V., "The effect of testosterone propionate on the genital tract of the immature female rat," *Endocrinology,* 23, 1938, pp.779-783.

Satyanarayana, K., J.Y. Bradfield, G. Bhaskaran, and K.H. Dahm, "Stimulation of vitellogenin production by methoprene in prepupae and pupae of Manduca sexta," *Arch Insect Biochem Physiol*, 25, 1994, pp.21-37.

Schreiber, G. and S.J. Richardson, "The evolution of gene expression, structure and function of transthyretin," *Comp Biochem Physiol B Biochem Mol Biol*, 116, 1997, pp.137-160.

Schwartz, H.L., J.H. Oppenheimer, and H.H. Samuels (Eds.), *Molecular Basis of Thyroid Hormone Action*, 1983, pp. 413-444.

Shaban, M., and P. Terranova, "2-Bromo-I-ergocryptine mesylate (CB-154) inhibits prolactin and luteinizing hormone secretion in the prepubertal female rat," *Biology of Reproduction*, 34, 1986, pp. 788-795.

Shanker, G., A.T. Campagnoni, and R.A. Pieringer, "Investigations on myclinogenesis *in vivo*; Developmental expression of myelin basic protein mRNA and its regulation by thyroid hormone in primary cerebral cell cultures from embryonic mise," *J. Neurosci. Res*., 17, 1987, pp. 220-224.

Shupnik, M.A., and E.C. Ridgway, "Thyroid hormone control of thyrotropin gene expression in rat anterior pituitary cells," *Endocrinol*, 121, 1987, pp. 619-624.

Shurin, J.B., and S.I. Dodson, "Sublethal toxic effects of cyanobacteria on nonylphenol on environmental sex determination and development in *Daphnia*," *Environ. Toxicol. Chem.,* 16, 1997, pp. 1269-1276.

Simental J.A., M. Sar, M.V. Lane, F.S. French, E.M. Wilson, *Journal of Biol Chem*, 266(1), 1991, pp. 510-518.

Song, M.K.H., B. Dozin, D. Grieco, J.E. Rall, and V.M. Nikodem, "Transcriptional activation and stabilization of malic enzyme mRNA precursor by thyroid hormone," *J. Biol. Chem*., 263, 1986, pp. 17970-17974.

Strait, K.A., H.L. Schwartz, A. Perez-Castillo, and J.H. Oppenheimer, "Relationship of c-erbA content to tissue triiodothyronine nuclear binding capacity and function in developing and adult rats," *J. Boil. Chem*., 265, 1990, pp. 10514-10521.

Summer, C.L., J.P. Giesy, S.J. Bursian, J.A. Render, T.J. Kubiak, P.D. Jones, D.A. Verbrugge, and R.J. Aulerich, "Effects Induced by Feeding Organochlorine-Contaminated Carp From

Saginaw Bay, Lake Huron, To Laying White Leghorn Hens. II. Embryonic and Teratogenic Effects," *J. Toxicol. Environ. Health*, 49, 1996, pp. 409-438.

Szelei, J., J. Jimenez, A.M. Soto, M.F. Luizzi, and C. Sonnenschein, "Androgen-induced inhibition of proliferation in human breast cancer MCF7 Cells Transfected with Androgen Receptor," *Endocrinology*, 138, 1997, pp. 1406-1412.

Tata, J.R., "Hormonal regulation of programmed cell death during amphibian metamorphosis," *Biochem Cell Biol*, 72, 1994, pp. 581-588.

Taton, M., et al., "Interaction of Triazole fungicides and plant growth regulators with microsomal cytochrome P-450-dependent obtusifoliol 14a-Methyl Demethylase," *Pestic Biochem. Physiol*., 90, 1988, pp. 1363-1370.

Taylor, T., F.E. Wondisford, T. Blaine, and B.D. Weintraub, "The paraventricular nucleus of the hypothalamus has a major role in thyroid hormone feedback regulation of thyrotropin synthesis and secretion," *Endocrinol*., 126, 1990, pp. 317-324.

Thompson, C.C., "Thyroid hormone-responsive genes in developing cerebellum include a novel synaptotagmin and a hairless homolog," *J. Neurosci*., 16, 1996, pp. 7832-7840.

Tilson, H.A., J.L. Jacobson, and W.J. Rogan, "Polychlorinated Biphenyls and the Developing Nervous System: Cross-Species Comparisons," *Neurotox Teratol.,* 12, 1990, pp. 239-248.

Timiras, P.S., and E.U. Nzekwe, "Thyroid hormones and nervous system development," *Biol Neonate*, 55, 1989, pp. 376-385.

Tori, G.M., and L.P. Mayer, "Effects of Polychlorinated Biphenyls on the Metabolic Rates of Mourning Doves Exposed to Low Ambient Temperatures," *Bull. Environ. Contam. Toxicol*. 27, 1981, pp. 678-682.

Tsai, S.Y., and M.H. Tsai, "Chick ovalbumin upstream promoter-transcription factors (COUP-TFs): coming of age," *Endocr Rev*, 18, 1997, pp. 229-240.

Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Test Methods. NIEHS, Final Report, March, 1997.

VanderKuur, J.A., M.S. Hafner, J.K. Christman, and S.C. Brooks, "Effects of estradiol-17beta analogues on activation of estrogen response element regulated chloramphenicol acetyltransferase expression," *Biochemistry*, 32, 1993b, pp. 7016-7021.

VanderKuur, J.A., T. Wiese, and S.C. Brooks, "Influence of estrogen structure on nuclear binding and progesterone receptor induction by the receptor complex," *Biochemistry*, 32, 1993, pp. 7002-7008.

Wakeling, A.E., B.J. Furr, A.T. Glen, and L.R. Hughes, *Journal of Steroid Biochemistry*, 15, 1981, pp. 355-359.

Waller, C.L., B.W. Juma, L.E. Gray Jr., and W.R. Kelce, "Three-dimensional quantitative structure-activity relationships for androgen receptor ligands," *Toxicology and Applied Pharmacology*, 137, 1996, pp. 219-227.

Wibbels, T., J.J. Bull, and D. Crews, "Synergism between temperature and estradiol: A common pathway in turtle sex determination?" *Journal of Experimental Zoology*, 260, 1991, pp.371-381.

Wondisford, F.E., J.A. Magner, and B.D. Weintraub, "Thyrotropin," in Braverman, L.E. and R.D. Utiger (Eds.), *Werner and Ingbar's The Thyroid: A Fundamental and Clinical Text*, 1996, pp. 190-206.

Yu, V.C., C. Delsert, B. Andersen, J.M. Holloway, O.V. Devary, A.M. Näär, S.Y. Kim, J.M. Boutin, C.K. Glass, and M.G. Rosenfeld, "RXRß: A coregulator that enhances binding of retinoic acid, thyroid hormone, and vitamin D receptors to their cognate response elements," *Cell*, 67, 1991, pp. 1251-1266.

Zacharewski, T.R., K.L. Bondy, P. McDonell, and Z.F. Wu, "Antiestrogenic effect of 2,3,7,8-tetrachlorodibenzo-p-dioxin on 17 beta-estradiol-induced pS2 expression," *Cancer Research*, 54, 1994, pp. 2707-13.

Zeiger, E., and W.S. Stokes, "Validating New Toxicology Tests for Regulatory Acceptance," *Regulatory Toxicology and Pharmacology*, 27, 1998, pp. 32-37.

Zhang, X., B. Hoffman, B.V. Tran, G. Graupner, and M. Pfahl, "Retinoid X receptor is anauxiliary protein for thyroid hormone and retinoic acid receptors," *Nature*, 335, 1992, pp. 441-446.

Zhou, M. Sar, J.A. Simental, M.V. Lane, and E.M. Wilson, "A ligand-dependent bipartite nuclear targeting signal in the human androgen receptor," *J. Biol. Chem*., 269, 1994, pp. 13115.

Zoeller, R.T., N. Kabeer, and H.E. Albers, "Molecular mechanisms of signal integration in hypothalamic neurons," *Amer. Zool*., 33, 1993, pp. 244-254.